Advanced Geoscience Targeting via Focused Machine Learning Applied to the QUEST Project Dataset, British Columbia

J. Granek, Geophysical Inversion Facility, University of British Columbia, Vancouver, BC, jgranek@eos.ubc.ca

E. Haber, Geophysical Inversion Facility, University of British Columbia, Vancouver, BC

Granek, J. and Haber, E. (2016): Advanced geoscience targeting via focused machine learning applied to the QUEST project dataset, British Columbia; *in* Geoscience BC Summary of Activities 2015, Geoscience BC, Report 2016-1, p. 117–126.

Introduction

As mineral exploration matures in the technological age, new methods for leveraging computational resources for prospecting continue to gain popularity. One such advance has been the widespread adoption of geographic information systems (GIS) for geoscientific-data management and integration, and the associated interest in spatial-data analysis. At the forefront of this effort is the emerging field of mineral-prospectivity mapping (MPM). With origins in Bayesian probability in the 1980s (Hart et al., 1978; Agterberg et al., 1990), the concept of linking numerous georeferenced datasets to derive a map of mineral potential has evolved to a suite of machine-learning algorithms, ranging from weights of evidence (WofE; Bonham-Carter et al., 1988; Agterberg et al., 1990; Carranza 2004) to neural networks (NN; Singer and Kouda, 1997; Porwal et al., 2003; Barnett and Williams, 2006) and support vector machines (SVM; Porwal et al., 2010; Zuo and Carranza, 2011; Abedi et al., 2012). Although many solution methods have been presented, certain characteristics of the MPM problem present difficulties related to uncertainty management and computational efficiency that have yet to be fully addressed. As a test case for this problem, the algorithm is run on the Geoscience BC QUEST (**Qu**esnellia **E**xploration **St**rategy) dataset that covers parts of NTS 093A, B, G, H, J, K, N and O, and comprises, amongst other available public data, airborne gravity, magnetics and electromagnetics, geochemical analysis, geological mapping and a database of known mineral occurrences in the region.

Mineral-Prospectivity Mapping

Mineral-prospectivity mapping (MPM) was first proposed in the late 1980s by geoscientists as a statistical method for the integration and interpretation of spatial patterns in geoscience data (Bonham-Carter et al., 1988). The concept was to determine the link between various geoscience datasets (i.e., geology, geophysics and geochemistry) and the existence or absence of economic mineralization (Figure 1).

The problem can be stated as follows: given training pairs of geoscience data $\mathbf{X}$, where the columns of $\mathbf{X}$ represent the different types of field measurements and the rows represent the different sample locations, and known mineralization occurrences $\mathbf{y}$ (binary indicator; mineralization or no mineralization), find some mapping function $f(\mathbf{X})$ that can approximate the relationship between the data and the mineralization occurrences such that it can be used to predict mineral potential on new data $\mathbf{X}^{new}$. In general, the mapping function $f(\sim)$ can be anything; Agterberg's original formulation, termed Weights of Evidence (Agterberg et al., 1990), used posterior probability as the mapping function. Numerous other approaches have been taken since, including fuzzy logic (Porwal et al., 2003), logistic regression (Harris and Pan, 1999), neural networks (Singer and Kouda, 1997; Barnett and Williams, 2006) and support vector machines (SVM; Zuo and Carranza, 2011; Abedi et al., 2012).

Due to the difficulty in field testing algorithms for this application and the relatively slow adoption of these methods by industry, most of the work—both past and present—on MPM has been primarily academic, although examples of governments and major mining companies using related methods do exist. That being said, some of the more commonly adopted methods include weights of evidence, fuzzy logic and neural networks. The popularity of these methods can be attributed to ease of use, flexibility and successful application in other fields. The primary challenge for this application is that geoscientific data are prone to errors, but none of the work mentioned above addresses this issue explicitly. This paper will do so by following the SVM approach and reformulating the problem as a total least squares optimization. In this way, the errors in the data can be faithfully and robustly treated.

Problem Characteristics

Although many black-box software packages exist for various machine-learning algorithms, the optimization problem presented by MPM has a number of practical character-

Keywords: *mineral-prospectivity mapping, machine learning, support vector machine, SVM, data mining, target generation, QUEST*

This publication is also available, free of charge, as colour digital files in Adobe Acrobat® PDF format from the Geoscience BC website: http://www.geosciencebc.com/s/DataReleases.asp.

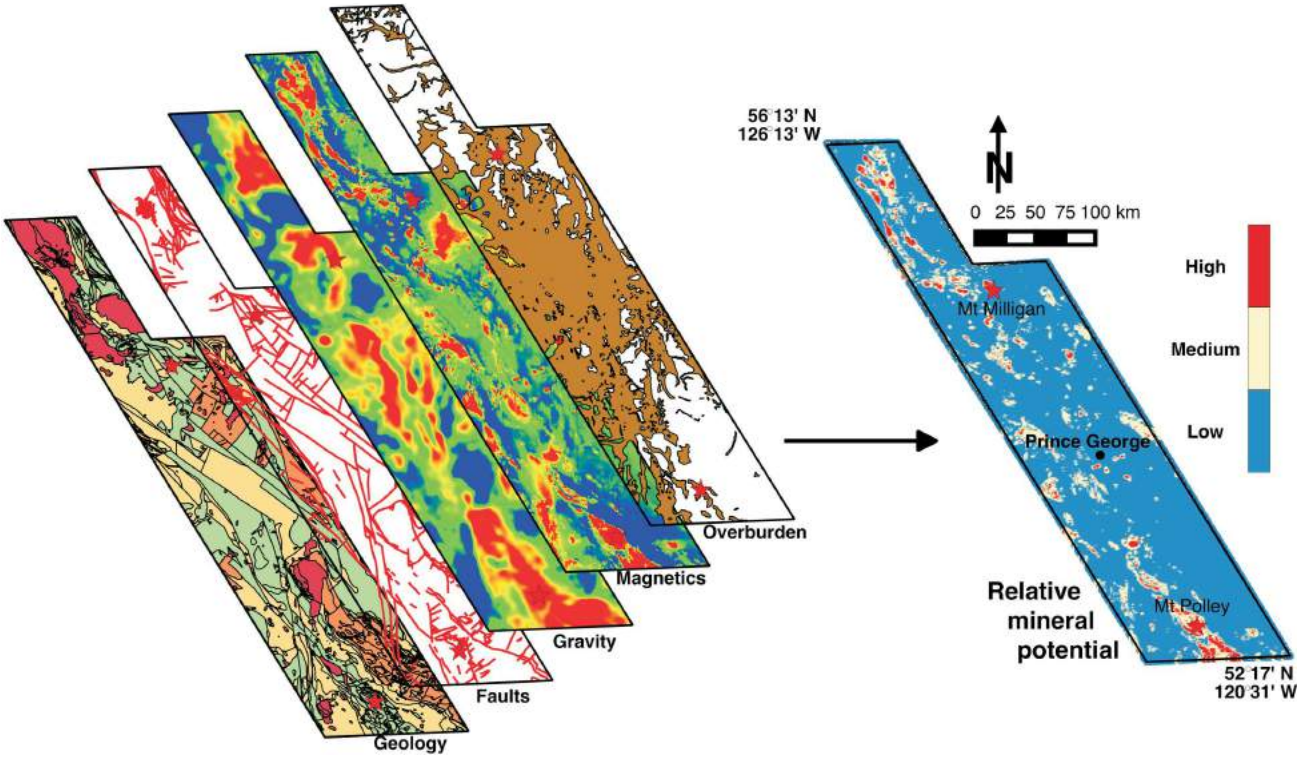

Figure 1. Combining numerous geoscientific datasets to produce a map of mineral relative potential using datasets from the QUEST study area (Barnett and Kowalczyk, 2007; Jackaman and Balfour, 2007).

istics that should be considered when implementing a solver. Many of these issues have been addressed individually by previous authors at one time or another; to date, however, a comprehensive solution for the MPM problem has yet to be discussed.

Imbalanced Training Data

In most supervised machine learning environments, an unbiased training is achieved by approximately sampling uniformly from each class. When this is not true, the problem is termed '**imbalanced**' (Chawla et al., 2004) and can lead to poor generalization of the resulting predictor. A number of methods exist to handle imbalanced data, including boosting (Guo and Viktor, 2004; Chawla et al., 2011) and rebalancing (Kubat and Matwin, 1997; Raskutti and Kowalczyk, 2004; Tang et al., 2009). As might be expected, mineral occurrences are relatively rare, resulting in an extremely imbalanced set of training labels. This problem is further exacerbated when one restricts the problem to a specific type of mineralization (e.g., porphyry deposits), as is often the case for prospectivity studies.

Training-Label Uncertainty

On top of the imbalanced nature of the MPM problem is the large degree of uncertainty associated with the training labels. In this regard, there are two fundamental problems:

- the crucial distinction that, in most cases, a label of 'no mineralization' simply means that mineralization has not been discovered, and not necessarily that there is none
- within each class (mineralized and not mineralized) exists a large range in certainty

For example, in many mineral occurrence databases, 'mineralization' encompasses occurrences ranging from producing mines all the way down to anomalies and prospects. Also, it is easy to understand how a classification of 'no mineralization' has very different implications in the middle of a highly explored mining district than it does in a remote location miles from the nearest field-sample site.

Training Data Uncertainty

As with any observed data, the training data in the MPM problem are associated with uncertainty from various sources. Some data, such as a magnetic-total-field measurement, will have numerical uncertainties associated with detection limits and processing procedures. Others, such as geological mapping of bedrock units, will have qualitative uncertainties associated with expert interpretation and sampling bias. Additionally, some data can have a spatially correlated uncertainty introduced by different exploration environments in the field (e.g., beneath thick deposits of overburden, it becomes prohibitively difficult to map bedrock). Unlike many machine-learning problems, where both the data and the labels can be trusted, it is known in mineral-prospectivity mapping that both have as-

sociated errors. Furthermore, the errors on the different data types can have very large statistical differences.

Big Data

Because mineral occurrences are rare events, the regions considered for MPM are often quite large. Combined with the large number of data *n* and the ever-growing number of predictive data sources (e.g., magnetic total field, fault locations, bedrock age), the data matrix **X** quickly becomes large and dense. Algorithms for solving this problem need to be able to handle large-scale learning of nonlinear relationships without prohibitive computational requirements.

Geoscience Datasets

Before one can contemplate how best to integrate numerous geoscience datasets, an understanding of the data is required. A typical exploration program will employ data from three primary disciplines: geology, geochemistry and geophysics. The variety of data within each of these is broad, comprising qualitative and quantitative measurements, inferred or interpreted values, and a large range in data resolution and uncertainty.

In an idealized exploration environment, all datasets would be densely sampled in the same locations, giving uniform coverage of the area of interest with an associated known uncertainty for each survey. The reality is more commonly represented by a scenario in which each survey was run independently, with highly varied sampling schemes, areas of coverage and target resolutions. The uncertainty on the data is typically a combination of numerous factors that also vary from survey to survey—many of which are either qualitative, inferred or simply unknown.

The integration of such surveys requires that the following considerations be properly handled:

- Data coverage may vary by survey, resulting in overlaps, holes or missing data.
- Nomenclature may vary amongst qualitative datasets describing the same parameters, resulting in either redundant or incompatible descriptions.
- Survey parameters may vary amongst quantitative datasets describing the same parameters, resulting in different sensitivities and measured values.
- Data quality and value can vary greatly, depending on such factors as age of the survey, sampling technique and environmental conditions.

Spatial Data

As one begins to collect various datasets for a given region, the first issue encountered is the discrepancy between the two data coverage plots shown in Figure 2. In most exploration environments, the available data will span multiple exploration programs, each potentially having a different area of interest. Even within a single exploration program, each survey will likely have a different sampling scheme based on the specific parameter being measured (e.g., bedrock geology will often be sampled wherever outcrop is available, whereas geophysical data are typically collected at predetermined grid points).

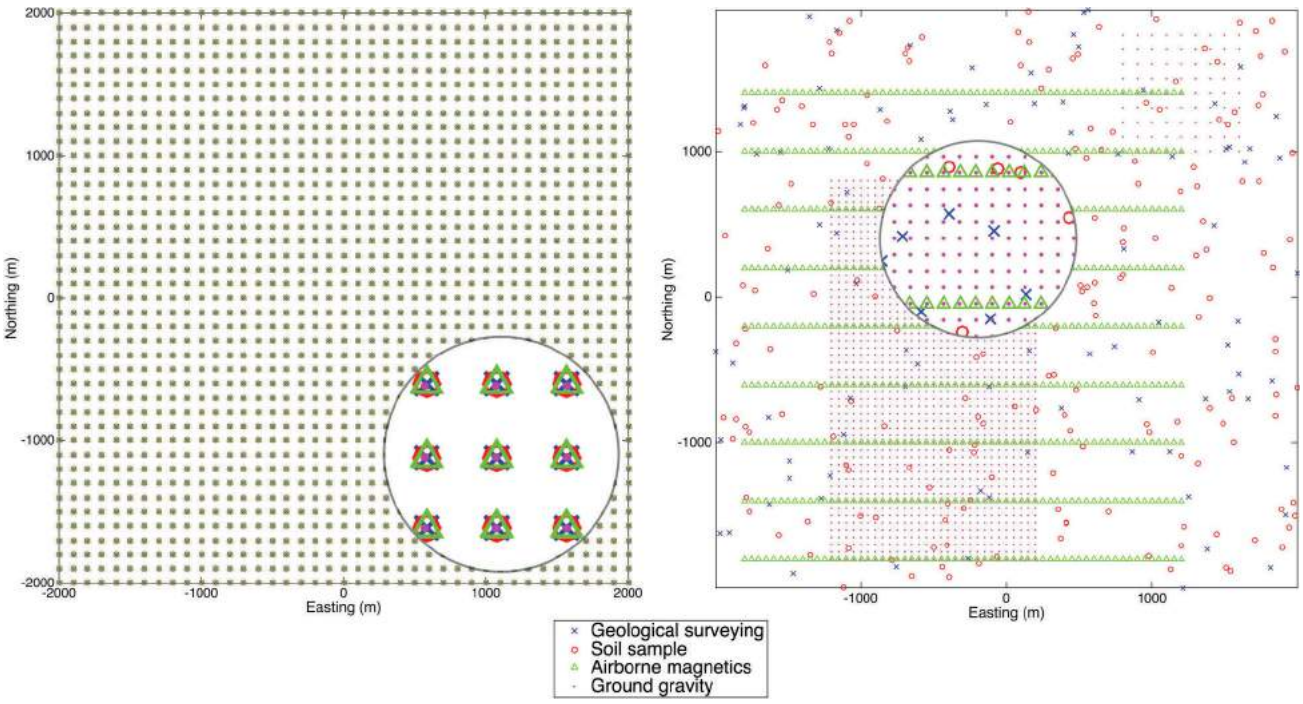

Figure 2. Idealized versus realistic data coverage for a representative suite of geoscience datasets.

To train the learning algorithm, the different measurements must have some geographic basis and, therefore, a domain must be specified. Traditional methods of prospectivity mapping, such as weights of evidence, handled this by reducing all field measurements—be they point measurements or polygons—to a series of overlapping polygons, each with a single value for each parameter (Agterberg et al., 1990).

The approach used in this project is to define a sample grid that specifies the target resolution of the prospectivity map. Continuous point measurements (i.e., geophysical fields or geochemical assays) can then be interpolated and resampled at the grid points, and discrete or categorical polygon layers (i.e., geological units) can simply be sampled at the specified locations. In this way, it is possible to define a spatial uncertainty for each data layer based on the distance from the field measurements to each grid point. Under this approach, sparse or missing data are easily handled as well, since they will simply have very large uncertainty, thus effectively removing their impact on the training of the algorithm.

Levelling/Reinterpreting Data

The other major challenge in integrating datasets occurs when multiple independent surveys are measuring the same parameters (e.g., ground magnetics and airborne magnetics). In such cases, it is important that steps be taken to ensure that all measurements of a given parameter are represented on the same scale or dictionary. For continuous measurements, such as geophysical fields or geochemical assays, this is commonly referred to as levelling (Luyendyk, 1997). For discrete or categorical data, this might be more complex and involve interpreting and reassigning labels to avoid redundant or incompatible descriptions (e.g., [volcanic rocks, intrusive rocks,…, dioritic intrusive rocks] → [volcanic rocks, intrusive rocks,…, intrusive rocks]).

Data Representation

For much of the continuous data (geophysics and geochemistry), representation simply consists of gridding the data and extracting values at points of interest. For some data, however, the question of data representation is more delicate. Take, for example, the bedrock class: this is a categorical variable where only one of the possible options can be true (sedimentary, volcanic, metamorphic or intrusive). If these data are represented as a single input with values ranging from 1 to 4 (mapping each of the four rock classes to a number), an inherent relationship is implied that states that sedimentary rocks are more similar to volcanic rocks than they are to intrusive rocks (since 1 is closer to 2 than it is to 4). This is not the desired behaviour. Instead, an approach has been taken to convert each of the classes to a separate binary input that states whether it is the specified rock type or not.

Another type of data that can be difficult to properly represent is point or line data (e.g., faults or mineralization occurrences). Because of the grid-sampling approach taken, data that are only defined on a point or a line can easily be misrepresented (if the data support is between the grid points) or simply saturated by the majority of the grid points that have no value. For these types of data, the approach taken is to define a range of influence or some other spatial function, such as distance to a fault or density of faults. This is done using different point-spread functions (e.g., Gaussians) to extend the support of these data from a single point or line to a continuous variable that can be sampled on the grid.

Support Vector Machines

Support vector machines (SVMs) were first formally introduced in the 1990s by Vapnik, Boser and Guyon (Boser et al., 1992; Vapnik, 2013) as a machine-learning algorithm structured on the statistical-learning theory (VC theory) developed by Vapnik and Chervonenkis during the 1960s and 1970s (Vapnik and Chervonenkis, 1971). The basic principle of SVMs is to construct an optimal margin classifier that has complexity based not on the dimensionality of the feature space, but rather on the number of support vectors, thus allowing for sparse solutions in high dimensions (Figure 3). The SVM is an optimal margin classifier because the goal is to learn the equation for a hyperplane that separates the different data classes with as large a margin as possible (subject to the constraints). A classifier with complexity independent of the data dimensionality can be a powerful advantage when dealing with large, highly dimensional datasets, since only a small subset is necessary in constructing the SVM classifier.

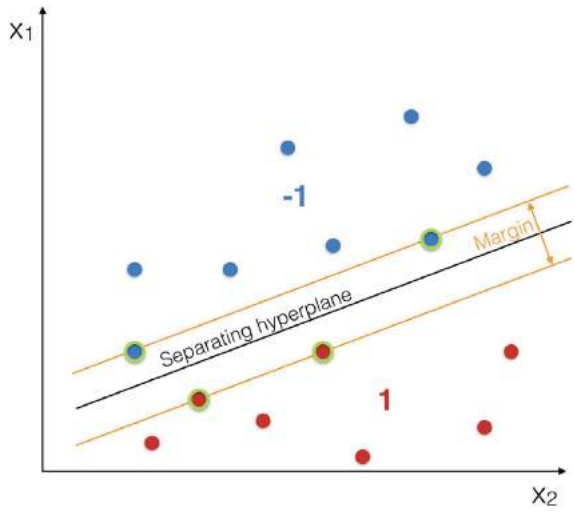

Figure 3. Optimal margin classification using the support vector machine (SVM); the separating hyperplane (black) is determined by maximizing the margin (yellow) between a few sparse support vectors (outlined in green).

The SVM falls under the branch of machine learning known as supervised learning, in which a predictor is taught using training data and training labels (Friedman et al., 2001). In the simplest case, one can consider training data with binary labels. The SVM can also easily be modified to handle regression on continuous or multivalued labels—binary classification is simply the most intuitive to visualize and illustrate the algorithm. The data and labels can therefore be represented as

$$[(\mathbf{X}_1, \mathbf{y}_1), (\mathbf{X}_2, \mathbf{y}_2),\ldots, (\mathbf{X}_n, \mathbf{y}_n)] \text{ with } \mathbf{X}_i R^m, \boldsymbol{y}_i \{-1, 1\}$$

For mineral-prospectivity mapping, $\mathbf{X}_i$ would be a vector of different field measurements (e.g., magnetic total field, bedrock age, distance to fault) for a given sample location and $\mathbf{y}_i$ would signify 'mineralization' or 'no mineralization' for that location. If the data are linearly separable, then one can define a separating hyperplane:

$$f(\mathbf{X}) = \mathbf{X}\mathbf{w} + \mathbf{b}$$

where $\mathbf{w}$ and $\mathbf{b}$ are weights with normalization $|\mathbf{X}_n\mathbf{w} + \mathbf{b}| = 1$. The problem then becomes one of maximizing the margin between training points of opposing classes –1 and 1 (see Figure 3). This is equivalent to the following optimization problem (Burges, 1998; Friedman et al., 2001):

$$\begin{aligned} &\underset{w}{\text{minimize}} && \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\ &\text{subject to} && \boldsymbol{y}_i(\boldsymbol{X}_i\boldsymbol{w} + \boldsymbol{b}) \geq 1, i = 1, \ldots, n \end{aligned}$$

which can be solved using quadratic programming (Burges, 1998; Platt, 1998; Smola and Schölkopf, 2004; Chang and Lin, 2011) or a number of iterative gradient-based methods (Joachims, 2006; Chapelle, 2007; Chang et al., 2008; Shalev-Shwartz et al., 2011). To solve this in the primal form using gradient-based methods, one typically casts the problem as a regularized risk minimization with unknown regularization parameter and iterates through perturbations of the model weights $\mathbf{w}$

$$\underset{w}{\text{minimize}} \quad \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2 + \boldsymbol{e}^T\max(0, 1 - \text{diag}(\boldsymbol{y})(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{b}))$$

where $\mathbf{e}$ is a vector of ones. This formulation is preferred not only due to its simplicity, but also because it is scalable with number of data n. Traditionally, however, the SVM problem was solved via quadratic programming using the primal-dual optimality conditions (KKT)

$$\underset{w}{\text{minimize}} \quad \frac{1}{2}\alpha^T\boldsymbol{Q}\alpha + (\mathbb{I}^T + \beta\boldsymbol{y}^T)\alpha$$

where $\mathbf{Q} = \text{diag}(\mathbf{y})\mathbf{X}^T\mathbf{X}\text{diag}(\mathbf{y})$. The primal and dual formulations were equally popular, however, until the dual formulation recently became more prevalent due to the natural extension to nonlinear transformations via the kernel trick (Figure 4; Boser et al., 1992). Since the data only appear as inner products in the dual form, a transformation $K(\mathbf{X},\mathbf{X}^T)$ can be applied and yet the size of K remains $n \times n$. This has powerful implications, allowing for transformations to very high dimensional spaces without penalty, but it is easy to see that this becomes computationally expensive when n is large.

This fact, combined with the realization that kernel methods can be applied in the primal form via the representer theorem (Chapelle 2007), has led to a resurgence of primal SVM methods in the past 10 years. With the advance of iterative optimization techniques, primal SVM methods have gained in speed and popularity to accommodate the growing size of applied datasets. Extensions of the algorithm to many specific problems have also taken place, including imbalanced datasets (Raskutti and Kowalczyk, 2004; Tang et al., 2009; Pant et al., 2011), density estimation (Chen et al., 2001; Manevitz and Yousef, 2002; Mordelet and Vert, 2010) and the incorporation of uncertainties for robust estimation (Zhang, 2005; Carrizosa, 2007; Pant et al., 2011; Huang et al., 2012).

SVM for Prospectivity Mapping

Beginning from the primal SVM formulation previously discussed, it is straightforward to adapt the problem to handle the characteristics specific to mineral-prospectivity mapping, namely by adding uncertainties on the data and

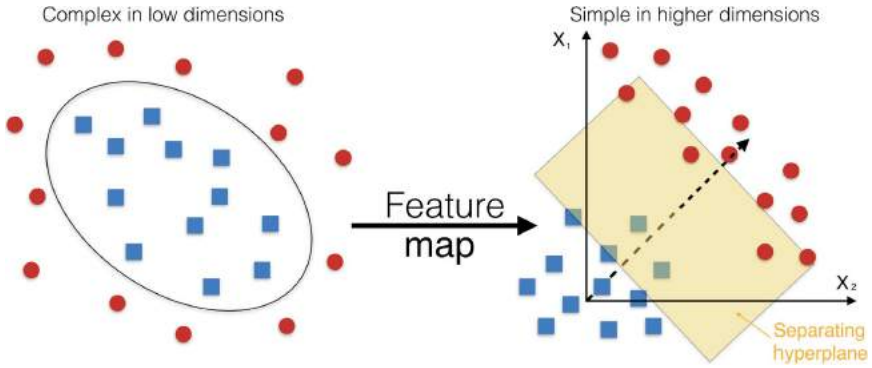

Figure 4. Nonlinear mapping via the kernel trick; classes may not be linearly separable in original dataspace, but easily separable in the transformed dataspace.

labels. In the simplest case, this can be done by adding weighting terms that penalize highly uncertain values, although more sophisticated algorithms can also be developed, such as in Granek and Haber (2015). In the simplest case (linearly separable SVM classification), training label uncertainty can be incorporated with the addition of a weighting term in the objective function

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2 + \left(\frac{1}{\boldsymbol{\varepsilon}}\right)^T \max\left(0, 1 - \text{diag}(\boldsymbol{y})(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{b})\right)$$

where ε_i is proportional to the uncertainty on the i^{th} training label. In this way, labels with low confidence are used less to train the classifier because they do not contribute as much to the misfit. Similarly, uncertainties in the training data can be handled by adding an extra penalty term (in red) to the objective function

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2 + \left(\frac{1}{\boldsymbol{\varepsilon}}\right)^T \max\left(0, 1 - \text{diag}(\boldsymbol{y})(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{b})\right) + \frac{\beta}{2}(\boldsymbol{X} - \boldsymbol{X}_{obs})^T \text{diag}\left(\frac{1}{\Sigma}\right)(\boldsymbol{X} - \boldsymbol{X}_{obs})$$

It is now assumed that the observed data $\mathbf{X}_{obs}$ are incorrect, and the true data are $\mathbf{X}$. Treating the uncertainties (Σ) as a proxy for the variance of the difference between these two data, and assuming Gaussian distribution, this bares a striking resemblance to the total least squares (Golub and Loan, 1980) problem.

Results

Finally, to demonstrate the utility of such an algorithm for mineral-prospectivity mapping (MPM), the following example from the QUEST (**Qu**esnellia **E**xploration **St**rategy) project in central British Columbia is presented. This work was funded by NEXT Exploration Inc. and Mitacs, so the results are currently confidential. For demonstration purposes, permission has been given to show preliminary results from a subset of the QUEST area (Figures 5, 6). This region is known to host a number of large, economic, copper-porphyry deposits. Through a government-sponsored program, a large amount of geoscientific data (including geological mapping, geochemical analysis and geophysical-data acquisition) was acquired between 2008 and 2012 in order to stimulate mineral exploration. Included in the large multidimensional dataset are covariates such as total-magnetic-field intensity, age of bedrock, geological rock class and copper content. Since each dataset was collected independently with its own sampling scheme, all layers were resampled to a base grid of 300 × 300 m, resulting in

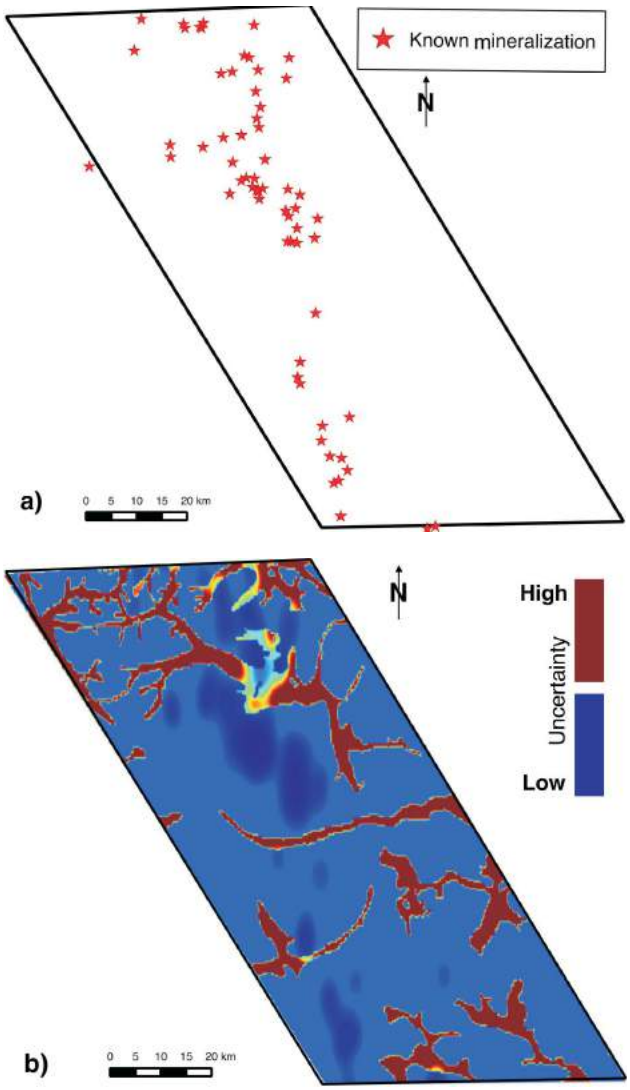

Figure 5. Examples of input layers for a subset of data from the QUEST project area: **a)** known mineralization locations, and **b)** uncertainty estimates for the mineralization labels (red is more uncertain). NEXT Exploration Inc. is thanked for permission to publish these preliminary results.

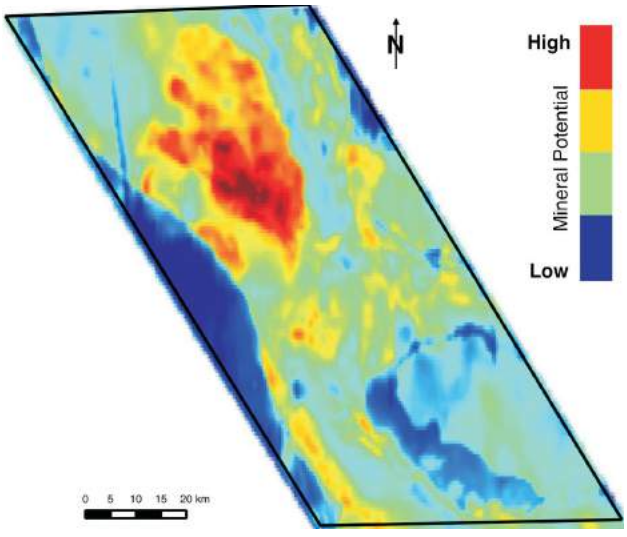

Figure 6. Prediction for prospective mineral regions in a portion of the QUEST area (red is more prospective, blue is less). NEXT Exploration Inc. is thanked for permission to publish these preliminary results; final prospectivity maps and locations are confidential under the terms of the funding agreement with NEXT Exploration Inc.

more than 700 000 sample points. When all data were assembled and properly processed for training, 91 distinct input layers were used, including both continuous and discrete values.

Uncertainty on these inputs can vary widely, depending on the data source. For example, most geophysical data can bear uncertainty in the form of a noise floor plus an acceptable standard deviation, whereas it is less obvious for geological data due to the subjective, interpreted nature of the measurements. In these cases, estimates can still be made based on confidence in the expert and the availability of field measurements.

As previously mentioned, the labels for the MPM problem present a suite of practical issues. The lack of confident negative labels (no mineralization) results in an imbalanced learning problem, and the sparse subjective nature of the positive labels (mineralization) results in a large range in confidence that can be adequately quantified using a framework of uncertainty estimates. For the QUEST dataset, 155 alkalic copper-porphyry–style mineral occurrences were used to generate a set of binary labels on the base grid. Each occurrence has associated with it a status ranging from 'Showing' to 'Producer' (six unique statuses are possible), indicating the confidence in the mineral occurrence being economic. Combining this with other factors such as the extent of the overburden (concealing potentially mineral-bearing bedrock), uncertainty estimates for the labels (see Figure 5b) were generated, ranging from 1 (confident label) to 50 (not confident label). The final result is a predicted mineral-prospectivity map (Figure 6) that indicates which regions are more favourable for copper-porphyry mineralization than others. As one can see, the algorithm was able to successfully predict the known prospective regions, as well as highlight potential new areas for exploration. The addition of uncertainty estimates in the algorithm provides a more robust framework for the incorporation of multidisciplinary data that possess a large range in data quality.

Discussion and Future Work

In the mineral-prospectivity–mapping literature, much has been made of the difference between knowledge-driven and data-driven methods. Although SVM automates the learning process and is therefore considered a data-driven approach, it is important to acknowledge that, without properly informed data processing prior to learning, the entire methodology is doomed to poor performance. This is not a short-coming of SVM but rather a well-known idiom of machine learning in general: "garbage in, garbage out!" In this sense, it is advisable to employ a data-driven approach, such as SVM, but with deliberate consultation of field experts when processing the data prior to learning.

A major shortcoming of many prospectivity-mapping methods, including SVM, is the insensitivity to spatial patterns in the data. One can think of the example of a typical porphyry halo, in which the presence of a large ring in the geophysical data is indicative of a target in the centre. In such a case, the actual value at each point location is not nearly as important as the structure it represents.

One promising new algorithm designed to handle the challenge is the convolutional neural network (CNN; LeCun and Bengio, 1995). An extension of the well-known neural network family of learning algorithms, CNNs implicitly build in sensitivity to spatial structures via use of convolutional kernels that are optimized to detect key structures in the input data. Developed and popularized in the last decade primarily for image, text and speech recognition, CNNs present an exciting new algorithm ideally suited for recognizing complex patterns in spatial geoscience datasets.

Conclusions

This work has explored the rapidly developing field of mineral-prospectivity mapping from an algorithmic and data-processing point of view. Although much work has been done in the fields of machine learning, geoscience and GIS independently, the intersection of all three has left a suite of challenges that remain to be fully addressed. Using a subset of the data from Geoscience BC's QUEST project, this study addresses several challenges through the modification and application of a primal support-vector-machine algorithm, incorporating uncertainties in both the labels and the data. In conjunction with this work, much effort has been directed at the proper understanding and use of data processing prior to learning. Going forward, the authors are exploring different algorithms that are better able to handle the spatial and structured nature of many geoscience datasets; most notably, current research is working on developing a convolutional neural network for mineral-prospectivity mapping. Such algorithms show promise as tools for geoscientists, both in the early stages of exploration by allowing for critical appraisal of which datasets are most valuable, and in the later stages to extract maximum value from existing data-rich environments.

Acknowledgments

Funding for this research comes from NEXT Exploration Inc. and Mitacs, as well as Geoscience BC. Thanks to P. Kowalczyk for reviewing the manuscript and providing helpful comments.

References

Abedi, M., Norouzi, G.-H. and Bahroudi, A. (2012): Support vector machine for multi-classification of mineral prospectivity areas; Computational Geoscience, v. 46, p. 272–283.

Agterberg, F., Bonham-Carter, G. and Wright, D. (1990). Statistical pattern integration for mineral exploration; *in* Computer Applications in Resource Estimation Prediction and Assessment for Metals and Petroleum, G. Gaál and D.F. Merriam (ed.), Pergamon Press, Toronto, p. 1–21.

Barnett, C.T. and Kowalczyk, P.L. (2007): Airborne electromagnetics and airborne gravity in the QUEST project area, Williams Lake to Mackenzie, British Columbia; *in* Geoscience BC Summary of Activities 2007, p. 1–6.

Barnett, C. and Williams, P. (2006): Mineral exploration using modern data mining techniques; First Break, v. 24, no. 7, p. 295–310.

Bonham-Carter, G., Agterberg, F. and Wright, D. (1989): Integration of geological datasets for gold exploration in Nova Scotia; Photogrammetric Engineering and Remote Sensing, v. 54, no. 11, p. 1585–1592.

Boser, B., Guyon, I. and Vapnik, V. (1992): A training algorithm for optimal margin classifiers; Proceedings of the Fifth Annual Workshop on Computational Learning Theory, p. 144–152.

Bottou, L. and Lin, C. (2007): Support vector machine solvers; *in* Large Scale Kernel Machines, L. Bottou, L. Chapelle, D. DeCoste and J. Weston (ed.), MIT Press, Boston, p. 301–320.

Burges, C.J.C. (1998): A tutorial on support vector machines for pattern recognition; Data Mining Knowledge Discovery, v. 2, p. 121–167, URL <http://research.microsoft.com/pubs/67119/svmtutorial.pdf> [November 2015].

Carranza, E.J.M. (2004): Weights of evidence modeling of mineral potential: a case study using small number of prospects, Abra, Philippines; Natural Resource Reserves, v. 13, no. 3, p. 173–187.

Carrizosa, E. and Plastria, F. (2007): Support vector regression for imprecise data; Vrije University, MOSI Working Paper 35, 35 p.

Chang, C.-C. and Lin, C.-J. (2011): LIBSVM: a library for support vector machines; ACM Transactions on Intelligent Systems and Technology, v. 2, no. 27, p. 1–27.

Chang, K.-W., Hsieh, C.-J. and Lin, C.-J. (2008): Coordinate descent method for large-scale L2-loss linear support vector machines; Journal of Machine Learning, v. 9, p. 1369–1398.

Chapelle, O. (2007) Training a support vector machine in the primal; Neural Computation, v. 19, no. 5, p. 1155–1178.

Chawla, N., Bowyer, K., Hall, L. and Kegelmeyer, W. (2011): SMOTE: synthetic minority over- sampling technique; Journal of Artificial Intelligence Research, v. 16, p. 321–357.

Chawla, N., Japkowicz, N. and Kotcz, A. (2004): Editorial: special issue on learning from imbalanced data sets; SIGKDD Explorations, v. 6, no. 1, p. 1–6.

Chen, Y., Zhou, X. and Huang, T. (2001): One-class SVM for learning in image retrieval; *in* Proceedings of the International Conference on Image Processing, v. 1, p. 34–37.

Friedman, J., Hastie, T. and Tibshirani, R. (2001): The Elements of Statistical Learning; Springer, Berlin, Springer Series in Statistics, Volume 1.

Golub, G. and Loan, C.V. (1980): An analysis of the total least squares problem; SIAM Journal Numerical Analysis, v. 17, no. 6, p. 883–893.

Guo, H. and Viktor, H. (2004): Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach; SIGKDD Explorations, v. 6, no. 1, p. 30–39.

Granek, J. and Haber, E. (2015): Data mining for real mining: a robust algorithm for prospectivity mapping with uncertainties; *in* Proceedings of the 2015 SIAM International Conference on Data Mining, p. 145–153. doi:10.1137/1.9781611974010.17

Harris, D. and Pan, G. (1999): Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis; Natural Resource Research., v. 8, no. 2, p. 93–109.

Hart, P., Duda, R. and Einaudi, M. (1978): PROSPECTOR—a computer-based consultation system for mineral exploration; Mathematical Geology, v. 10, no. 5, p. 589–610. doi:10.1007/bf02461988

Huang, G., Song, S., Wu, C. and You, K. (2012): Robust support vector regression for uncertain input and output data; IEEE Transactions on Neural Networks and Learning Systems v. 23, no. 11, p. 1690–1700.

Jackaman, W. and Balfour, J.S. (2008): QUEST project geochemistry: field surveys and data re-analysis (parts of NTS 93A, B, G, H, J, K, N and O); *in* Geoscience BC Summary of Activities 2007, Geoscience BC, Report 2008-1, p. 7–10.

Joachims, T. (2006): Training linear SVMs in linear time; *in* Proceedings of the 12th ACM SIGKDD International Conference Knowledge Discovery and Data Mining (KDD), 10 p.

Kubat, M. and Matwin, S. (1997): Addressing the curse of imbalanced training sets: one-sided selection; *in* Proceedings of the Fourteenth International Conference on Machine Learning, p. 179–186.

LeCun, Y. and Bengio, Y. (1995): Convolutional networks for images, speech, and time series; *in* The Handbook of Brain Theory and Neural Networks, M.A. Arbib (ed.), MIT Press, Boston, p. 255–258.

Luyendyk, A.P.J. (1997): Processing of airborne magnetic data; AGSO Journal of Australian Geology and Geophysics, v. 17, p. 31–38.

Manevitz, L. and Yousef, M. (2002): One-class SVMs for document classification; Journal of Machine Learning Research, v. 2, p. 139–154.

Mordelet, F. and Vert, J. (2010): A bagging SVM to learn from positive and unlabeled examples; Pattern Recognition Letters, v. 37, p. 201–209.

Pant, R., Trafalis, T. B. and Barker, K. (2011): Support vector machine classification of uncertain and imbalanced data using robust optimization; *in* Proceedings of the 15th World Scientific and Engineering Academy and Society (WSEAS) International Conference on Computers, p. 369–374.

Platt, J. (1998): Sequential minimal optimization: a fast algorithm for training support vector machines; Microsoft Technical Research, Technical Report MSR-TR-98-14, 21 p., URL <http://research.microsoft.com/pubs/69644/tr-98-14.pdf > [November 2015].

Porwal, A., Carranza, E. and Hale, M. (2003): Artificial neural networks for mineral-potential mapping: a case study from Aravalli Province, western India; Natural Resources Research, v. 12, no. 3, p. 155–171.

Porwal, A., Yu, L. and Gessner, K. (2010): SVM-based base-metal prospectivity modeling of the Aravalli Orogen, northwestern India; EGU General Assembly, Conference Abstracts, v. 12, p. 15171.

Raskutti, B. and Kowalczyk, A. (2004). Extreme re-balancing for SVMs: a case study; SIGKDD Explorations, v. 6, no. 1, p. 60–69.

Shalev-Shwartz, S., Singer, Y., Srebro, N. and Cotter, A. (2011): Pegasos: Primal Estimated sub-GrAdient SOlver for SVM; Mathematical Programming, v. 127, no. 1, p. 3–30.

Singer, D.A. and Kouda, R. (1997): Use of a neural network to integrate geoscience information in the classification of mineral deposits and occurrences; *in* Proceedings of Exploration 97, Fourth Decennial International Conference on Mineral Exploration, A.G. Gubins (ed.), p. 127–134.

Smola, A.J. and Schölkopf, B. (2004): A tutorial on support vector regression; Statistics and Computing, v. 14, no. 3, p. 199–222.

Tang, Y., Zhang, Y.Q., Chawla, N.V. and Krasser, S. (2009): SVMs modeling for highly imbalanced classification; Systems, Man and Cybernetics, Part B: Cybernetics, IEEE Transactions, v. 39, no. 1, p. 281–288.

Vapnik, V.N. and Chervonenkis, A.Y. (1971): On the uniform convergence of relative frequencies of events to their probabilities; Theory of Probability and its Applications, v. 16, no. 2, p. 264–280.

Vapnik, V. (2013): The Nature of Statistical Learning Theory; Springer Science and Business Media.

Zhang, J.B.T. (2005): Support vector classification with input data uncertainty; *in* Proceedings of Advances in Neural Information Processing Systems Conference, v. 17, p. 161–168.

Zuo, R. and Carranza, E.J.M. (2011): Support vector machine: a tool for mapping mineral prospectivity; Computers and Geoscience, v. 37, no. 12, p. 1967–1975.