Learning 3D Geological Structure from Drill-Rig Sensors for Automated Mining

Sildomar T. Monteiro[1], Joop van de Ven[2], Fabio Ramos[2] and Peter Hatherly[1]
[1,2]Australian Centre for Field Robotics
[1]School of Aerospace, Mechanical and Mechatronic Engineering
[2]School of Information Technologies
The University of Sydney, NSW, Australia
s.monteiro@acfr.usyd.edu.au

Abstract

This paper addresses one of the key components of the mining process: the geological prediction of natural resources from spatially distributed measurements. We present a novel approach combining undirected graphical models with ensemble classifiers to provide 3D geological models from multiple sensors installed in an autonomous drill rig. Drill sensor measurements used for drilling automation, known as measurement-while-drilling (MWD) data, have the potential to provide an estimate of the geological properties of the rocks being drilled. The proposed method maps MWD parameters to rock types while considering spatial relationships, i.e., associating measurements obtained from neighboring regions. We use a conditional random field with local information provided by boosted decision trees to jointly reason about the rock categories of neighboring measurements. To validate the approach, MWD data was collected from a drill rig operating at an iron ore mine. Graphical models of the 3D structure present in real data sets possess a high number of nodes, edges and cycles, making them intractable for exact inference. We provide a comparison of three approximate inference methods to calculate the most probable distribution of class labels. The empirical results demonstrate the benefits of spatial modeling through graphical models to improve classification performance.

1 Introduction

Mining is a multi-billion dollar business employing thousands of people worldwide. Given the current increasing demand for mineral commodities, such as iron ore, more efficient and automated processes are needed in the industry. The ultimate goal is to develop a fully autonomous, remotely operated mine. A major challenge for autonomous mining is to build accurate representations of the in-ground geology to determine the quantity and quality of the minerals of interest. Modern autonomous drill rigs are equipped with multiple sensors that provide measurements while drilling (MWD), which are normally used to monitor and control the drilling process. Characterizing subsurface geology from drilling measurements can be of substantial value for the mining industry. The accurate assessment of lithology and rock strength can be used to maximise the recovery of the desired rock types and improve blasting design by accurately determining the optimal explosive load and distribution.

This paper addresses the problem of relating MWD data to geotechnical properties of the rocks being drilled while taking into account the spatial information. We propose a novel approach that accounts for spatial relationships using conditional random fields (CRFs). CRFs are very powerful for modeling relational information, spatial relationships and other types of contextual information [Sutton and McCallum, 2007]. By directly modeling the conditional probability of the hidden states, given observations, rather than the joint probability, CRFs avoid the difficult task of specifying a sensor model for observations, as required by techniques such as hidden Markov models or Markov random fields. CRFs can, thus, handle arbitrary dependencies between observations which give them significant flexibility in modeling complex geological dependencies in the data.

For the problem of modeling geological structure, learning of CRF parameters can be done efficiently [Monteiro *et al.*, 2009]; in this paper we optimize the pseudolikelihood. However, inference of CRF models can pose a challenge, depending on the complexity of the graph structure. We provide a comparison of three message-passing-based inference algorithms on graphs with different degrees of complexity. We evaluate loopy belief propagation [Pearl, 1988] and two methods based on linear programming relaxation: sequential tree reweighted (TRW-S) [Kolmogorov, 2006] and max-product linear programming (MPLP) [Sontag *et al.*, 2008].

CRFs are a type of log-linear model for structured learning. The proposed method combines CRF models with boosting classifiers in order to obtain a nonlinear classification of the observations. Boosting is a machine learning technique for supervised classification that has a sound theoretical foundation and often yields accurate classification while being robust to overfitting [Hastie *et al.*, 2009]. The set of labels classified by boosting are used in the CRF model to learn local parameters discriminatively. In other words, boosting provides an initial estimate by combining the different sensor measurements and the CRF then improves this prediction to be spatially consistent. The resulting CRF model specifies the spatial relationship between MWD data providing an

improved rock classification of the subsurface geology.

The main contributions of this paper are: (1) proposing a novel approach to model 3D underground geological structure from multiple sensor measurements; (2) presenting an empirical comparison of three modern inference algorithms on a challenging real-world problem; (3) integrating two distinct machine learning techniques: undirected graphical models and boosting classifiers.

2 Related Work

The first attempts on relating drilling measurements to geotechnical properties of rocks focused on determining empirical indices as a proxy for rock strength, e.g. [Teale, 1965; Scoble *et al.*, 1989]. More recently, there have been a few studies applying machine learning techniques to process MWD data. Previous methods, such as the works of [Utt, 1999; Itakura *et al.*, 2004], mainly focused on using neural networks and their variants. The method presented in [Kadkhodaie-Ilkhchi *et al.*, 2010] used boosting classifiers to map drilling measurements to rock properties. However, none of those studies modeled the spatial dependencies of nearby geology.

The paper by [Monteiro *et al.*, 2009] attempts to use CRFs to model drilling measurements. However, they relied on a linear-chain CRF and could only associate measurements within individual drill holes. Our proposed method builds more complex graph structures and is, therefore, able to associate data of sections between neighboring drill holes as well. They used sum-product belief propagation for inference, whereas, in our method, the increased complexity of the resulting graph structure demanded the use of more sophisticated inference algorithms.

Our comparative study of different inference algorithms is somewhat similar to the one presented by [Szeliski *et al.*, 2008] in computer vision. However, in our study we included a more recent algorithm, max-product linear programming, instead of graph cuts, and we compare their performance in a different domain.

There are other related approaches that combine CRFs with boosted decision trees. In particular, gradient tree boosting has been successfully applied to learn CRF parameters [Dietterich *et al.*, 2004] and to train CRFs for logical sequences [Gutmann and Kersting, 2006]. Although learning CRF parameters using boosting is a possible alternative to pseudolikelihood, we chose to focus the paper on comparing inference methods instead of learning methods.

3 Conditional Random Fields

CRFs are discriminative, undirected graphical models that were originally proposed for labeling relational data [Lafferty *et al.*, 2001]. CRFs directly model $p(\mathbf{x}|\mathbf{z})$: the conditional distribution over the hidden variables $\mathbf{x}$ given observations $\mathbf{z}$, where $\mathbf{x} = \langle \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \rangle$, and $\mathbf{z} = \langle \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n \rangle$. The nodes $\mathbf{x}_i$, along with the connectivity structure represented by the undirected edges define a conditional distribution $p(\mathbf{x}|\mathbf{z})$ over the hidden states $\mathbf{x}$. The edges in the graph represent potential functions which map sensor measurements to non-negative numbers. By using log-linear combinations of potential functions where local potentials are denoted as $h(\mathbf{x}_i, \mathbf{z}_i)$ and pairwise potentials as $g(\mathbf{x}_i, \mathbf{x}_j)$, the conditional probability distribution is written as:

$$p(\mathbf{x} \mid \mathbf{z}) = \frac{1}{Z(\mathbf{z})} \exp \left\{ \sum_i \sum_{k=1}^{K_1} w_k^h h_k(\mathbf{z}_i, \mathbf{x}_i) + \sum_{i,j} \sum_{k=1}^{K_2} w_k^g g_k(\mathbf{x}_i, \mathbf{x}_j) \right\}, \tag{1}$$

where $\mathbf{w}^h$ is a vector with K_1 dimensions representing the weights for local potentials, $\mathbf{w}^g$ is a vector with K_2 dimensions representing the weights for the pairwise potentials and $Z(\mathbf{z})$ is a normalizing partition function. In our problem, the pairwise potential function associates neighboring nodes (borehole sections) in the graph while the local potential function associates nodes to observations (sensor measurements).

3.1 Inference

A CRF, together with its parameters, can be used to estimate the labels of new instances of unlabeled data. This step is referred to as inference. Inference in CRFs can estimate either the marginal distribution of each hidden variable $\mathbf{x}_i$ or the most likely configuration of all hidden variables $\mathbf{x}$ (*i.e.*, MAP estimation), as defined in (1). Both tasks can be solved using message passing algorithms, which works by sending local messages through the graph structure of the model [Koller and Friedman, 2009].

For graphs with no loops, such as chains or trees, it is possible to compute inference in closed form by message passing; this method is called belief propagation (BP) [Pearl, 1988]. For more complex graphical models, containing many loops, exact inference is not feasible. Furthermore, if the graph has a high number of nodes, exact inference can also rapidly become intractable. Our problem presents both characteristics, high number of nodes and loops, which make inference particularly challenging. Therefore, we resort to approximate inference techniques. We compare loopy BP with two promising recent algorithms.

Loopy belief propagation

Since message updates in BP are only local, the method can be easily extended to graphs with loops. The intuition is to propagate messages in the graph while minimizing the overall energy. The contributions from the loops diminish as the influence they cause into the graph reduces. Although optimality is not guaranteed, the resulting *loopy* belief propagation (LBP) often provides a good approximation to the solution in a range of applications, e.g. [Cho *et al.*, 2010]. Moreover, recent theoretical studies have provided some additional justification for applying LBP to graphs with cycles [Wainwright and Jordan, 2008].

Sequential tree reweighted message passing

The TRW max-product is a message passing algorithm somewhat similar to LBP [Wainwright *et al.*, 2005]. However, unlike LBP it has some convergence guarantees. It attempts to find the most probable configuration of undirected graphs based on a linear programming (LP) relaxation of an integer

program for the problem. TRW approximates a loopy graph by a convex combination of tree-structured graphs such as spanning trees. We use a variant of the TRW algorithm called sequential tree reweighted (TRW-S) [Kolmogorov, 2006], which has better convergence properties than previous versions. Although in TRW-S the lower bound estimate is guaranteed not to decrease, there is no stability guarantees for the energy itself, which may start to oscilate [Szeliski *et al.*, 2008].

Max-product linear programming
MPLP is a recent algorithm that can also be seen as an extension of the LP relaxation approach [Globerson and Jaakkola, 2008]. MPLP has similar convergence guarantees as TRW-S, derived from the properties of the LP relaxation. Although both MPLP and TRW-S are local methods not guaranteed to provide global convergence in general, they have shown to perform well in many practical applications in which standard LP relaxation or LBP methods have difficulty [Werner, 2010]. MPLP has the advantage of having no parameters to tune and present better results compared to TRW-S, although with higher computational cost. We use an extension of the MPLP, due to [Sontag *et al.*, 2008], that is able to obtain tighter relaxations by using a convex combination of clusters, which are calculated iteratively. This method approximates the true MAP problem and, if convergence is achieved, has global optimality guarantees.

3.2 Parameter Learning

The goal of CRF parameter learning is to determine the weights of the feature functions used in the conditional likelihood (1). CRFs can learn these weights discriminatively by maximizing the conditional likelihood of training data. Unfortunately, this optimization runs an inference procedure at each iteration, which is intractable in our case.

Therefore, we resort to maximizing the pseudolikelihood of the training data, which is given by the sum of local likelihoods $p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i))$, where $\mathrm{MB}(\mathbf{x}_i)$ is the Markov blanket of variable $\mathbf{x}_i$: the set of the immediate neighbors of $\mathbf{x}_i$ in the CRF graph [Besag, 1975]. Optimisation of this pseudolikelihood is performed by minimizing the negative of its log, resulting in the following objective function:

$$L(\mathbf{w}) = -\sum_{i=1}^{n} \log p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i), \mathbf{w}) + \frac{(\mathbf{w}-\widetilde{\mathbf{w}})^T(\mathbf{w}-\widetilde{\mathbf{w}})}{2\sigma^2}, \quad (2)$$

where the terms in the summation correspond to the negative pseudo log-likelihood and the right term represents a Gaussian shrinkage prior with variance σ^2. Without additional information, the prior mean is typically set to zero.

The gradient of the log of the pseudolikelihood can be computed extremely efficiently, without running an inference algorithm. We perform this optimization using unconstrained L-BFGS [Nocedal and Wright, 2000]. Learning by maximizing pseudolikelihood has been shown to perform very well in different domains, e.g. [He and Zemel, 2008].

4 CRFs for Modeling 3D Geological Structure

MWD data is typically logged sequentially, down the hole, during drilling. A straightforward approach is then to build a CRF model in the axial direction of the blast hole by using a chain-like structure, as illustrated in Fig. 1. The ith section of a blast hole composed of n sections is modeled as two interconnected nodes $\mathbf{z}_i$ and $\mathbf{x}_i$ representing the set of drilling measurements and the rock category, respectively. Drilling measurements are considered observed variables and are represented by shadowed nodes. Blast-hole section categories, which are not observed, correspond to latent (hidden) variables and are represented by clear nodes. The relationships between nearby blast-hole sections are represented by edges connecting them. However, the CRF chain only models within each hole and ignores the three-dimensional nature of the underground geology.

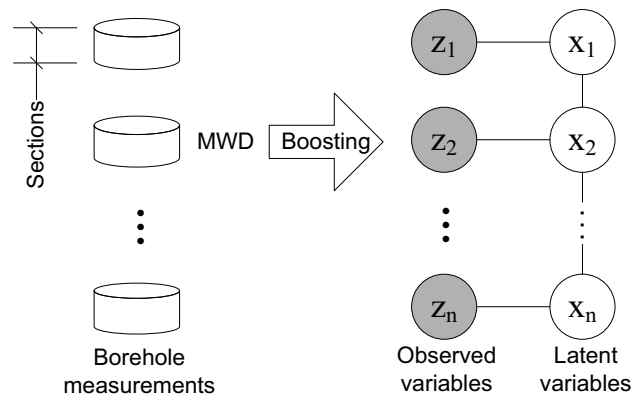


Figure 1: Graphical model of a CRF to model the spatial association between neighboring blast-hole sections. The observations z_i correspond to drilling measurements and the latent variables x_n indicate the corresponding classes.

4.1 Three dimensional graph structures

Our CRF approach attempts to model the 3D structure underground by considering relationships between neighboring holes. While inferring the optimal model structure is an intractable problem, we make a few assumptions to propose a practical method to build the 3D structures. We propose to model the problem as a 3D quasi-regular lattice-like graph.

We devised an algorithm that takes into account the spatial distribution of the holes and assign edges to the nearest hole aligned to one of the axes. Our basic assumption is that the actual distances between nodes remain fairly constant at each dimension (x, y and z axis). In this approach, edges associated with the same axis can be modeled by the same pairwise feature weight. If the data is perfect, i.e., it is distributed in a strict regular grid pattern and has no missing values, the algorithm will build a cubic structure as illustrated in Fig. 2, an example of 16 holes equally spaced containing four sections in each hole.

4.2 Local features

The CRF model can employ arbitrary feature functions to describe any particular property of the data. Instead of learning the CRF model directly from the raw observations, e.g. MWD data, it is sometimes advantageous to extract features from

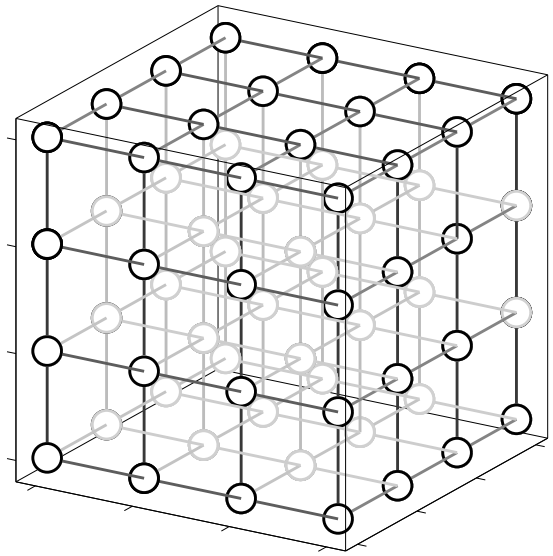


Figure 2: Graph representing the edge connections between latent variables of a cubic MWD data set; observation nodes and local edges are omitted. Pairwise edges having the same color indicate those that share the same weight value.

the data using a classification algorithm. This is because the CRF's log-linear representation, which corresponds to a unimodal Gaussian likelihood on each feature function, might not be flexible enough to model complex multimodal relationships. We use boosting classifiers to provide a nonlinear mapping from MWD data to rock categories, at each local section of a blast hole.

Boosted trees

The concept of boosting is to train many weak learners on various distributions of the input data and then combine the classifiers produced into a single committee [Hastie *et al.*, 2009]. Initially, the weights of all training examples are set equally, but after each round of the algorithm, the weights of incorrectly classified examples increase. The final committee, or ensemble, is a weighted majority combination of M weak classifiers and can be expressed as

$$h_k(\mathbf{z}) = sign\left(\sum_{m=1}^{M} \alpha_m^k C_m^k(\mathbf{z})\right), \tag{3}$$

where α_m quantifies the contribution of each respective weak classifier C_m.

We implemented a variant of the boosting algorithm called LogitBoost (LB) [Friedman *et al.*, 2000], which fits additive logistic regression models by stagewise optimisation of the maximum likelihood. It can be generalized to handle multiple classes by using a symmetric multiple logistic transformation. As weak learners, we used single-node decision trees, also known as regression stumps. The boosting scores (3) are used as local features.

4.3 Pairwise features

A pairwise feature is used to associate measurements from neighboring sections. The function associating a node $\mathbf{x}_i$ to a neighboring node $\mathbf{x}_j$ is defined as

$$g_k(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} a & \text{if } \mathbf{x}_i = \mathbf{x}_j \\ b & \text{if } \mathbf{x}_i \neq \mathbf{x}_j \end{cases} \tag{4}$$

where a and b are parameters that just have to be distinct to indicate equality or inequality. Those parameters need not be learned but can simply be assigned values taken from the dataset's training labels. Only the weights multiplying the indicator features are learned using pseudolikelihood. Note that for the three-dimensional graph, $K_2 = 3$ in (1) and, therefore, $k = \{1,2,3\}$ in (4), corresponding to the x, y and z axis, respectively. In other words, for each dimension, k, there is a different weight, w_k in (1), for the corresponding pairwise feature, g_k.

5 Experimental evaluation

The CRF method was evaluated using MWD data collected from a blast-hole drill rig, which is shown in Fig. 3. From a total of 17 drill sensors recorded, 5 measurements were manually selected for analysis: rotation speed, rotation pressure, pull-down rate, pull-down pressure, and bit air pressure; the excluded measurements are mainly binary flags indicative of the state of the drill's hydraulic system. Since each sensor has a different sampling rate, the measurements need to be resampled and grouped into appropriate sections of 10 cm depth intervals. MWD data is typically collected while drilling blast holes in the same vicinity on a "mine bench." We chose to present results of a representative bench that contains the three main rock types found in our study area, as the performance of the algorithm in other benches followed a similar trend. The chosen data set consists of 180 blast holes with an average depth of 14 m.

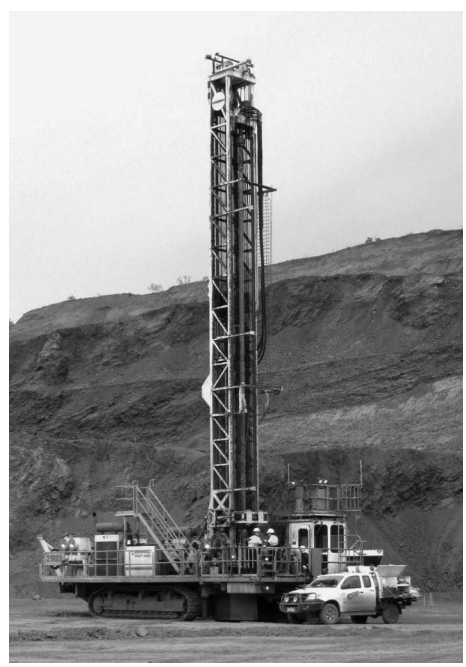


Figure 3: Drill-rig used to collect MWD data.

The ground-truth labels were determined by mine geologists using a combination of geophysical, chip and core logs. Note that this is a subjective process that creates minor uncertainty in the labels. A hierarchical labeling scheme was devised to group the geological zones into categories. The main geological categories are: shale, ore and banded iron formation (BIF). Each category can be further divided based on rock strength.

The numerical performance of the proposed method was evaluated by calculating accuracy, precision, recall, F-score, and area under the ROC curve (AUC); for details on those metrics, see [Sokolova and Lapalme, 2009]. The overall performance for all metrics except accuracy was calculated by *macro*-averaging. Accuracy was *micro*-averaged to avoid optimistic bias. The models were evaluated using 3-fold cross-validation. The cross-validation sets were not randomly sampled, but selected based on their spatial distribution to allow testing of the 3D structure of the CRFs. For comparison purposes, in all experiments the number of weak learners in the boosting algorithms was constant, 50. Nevertheless, the boosting algorithm is quite resilient to overfitting and we observed that using more weak learners does not degrade performance severely.

To investigate the effect of increasing the complexity of the graph, we examined four scenarios. We started with a graph without edges, which corresponds to solely applying LB classification, i.e., no pairwise features in the CRF, only the local features. Then we added the vertical edges (z axis), which corresponds to a chain-like structure; we refer to this graph as one-dimensional (1D) CRF. Next, we added the horizontal edges (x and y axis) one after the other; we refer to the resulting graphs as two-dimensional (2D) CRF and three-dimensional (3D) CRF, respectively. For the selected data set, the total number of nodes was $24,896$. An illustration of the graph structure for this data set is shown in Fig. 4. The total number of edges in the full 3D graph was $67,564$.

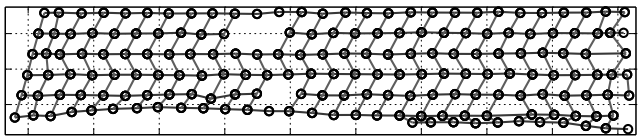


Figure 4: Birdseye view of the graph structure for the data set with 180 blast holes. Only the nodes at the top and the edges connecting them are shown.

The algorithms—CRF graph building, pseudolikelihood optimization and LBP—are implemented in MATLAB. Code for the TRW-S[1] and MPLP[2] is implemented in C/C++, as provided by the respective authors. An empirical comparison of the processing time for all three inference methods is presented in Fig. 5; the total processing time presented is the sum of all cross-validation folds. The experiments ran on a 16-core Linux computer. Note that the different times for LB reflect the basic computational overhead of each inference algorithm, which occur even when there is no edges in the graph, not differences in the performance of LB. This comparison was not intended to be a strictly fair comparison, since the implementations were not optimized for speed and LBP was in a different language. It can, however, provide a good indication of how each algorithm scale to more complex graph structures.

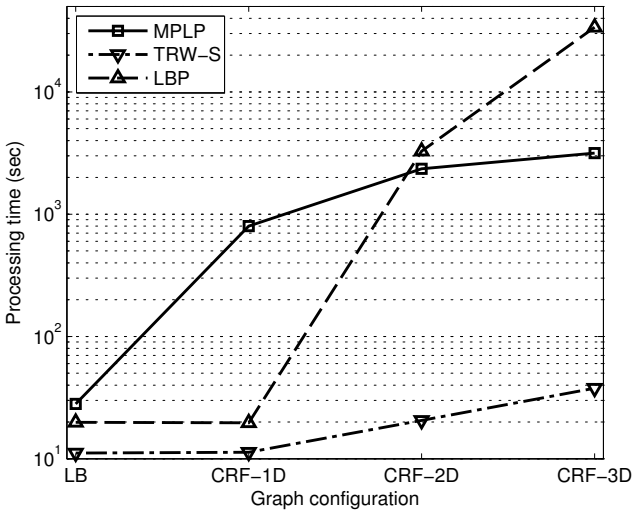


Figure 5: Experimental inference time required for the inference algorithms on different graph structure configurations. Note that the y-axis is in logarithmic scale.

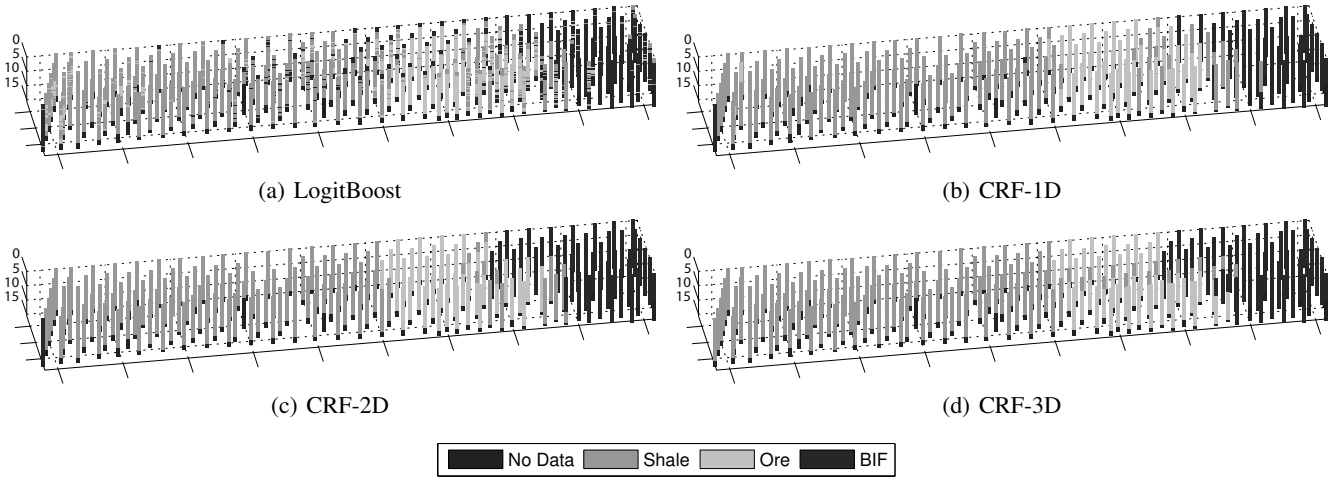
A quantitative analysis of the algorithms' performance on all 180 blast holes is presented in Table 1, for the 3 categories scheme. The qualitative results for estimating the geology of all 180 blast holes using the 3 categories scheme is presented in Fig. 6. Note that the experiments include a comparison with two previous methods, specifically, the LB (LogitBoost only) method, proposed in [Kadkhodaie-Ilkhchi *et al.*, 2010], and the 1D (chain CRF) method, proposed in [Monteiro *et al.*, 2009].

6 Discussion

The main goal of our proposed approach was to model the geology from multiple sensors installed in an autonomous drill rig. We proposed a method to map the geology in 3D by combining an undirected graphical model, CRFs, with a robust classifier, LogitBoost. The combined approach is able to reason about the spatial dependencies while performing nonlinear, multiclass classification. By modeling the spatial relationships we are able to exploit the physical assumption that, locally, the rock types are homogeneous. Performing inference on graphical models for this problem turned out to be very difficult since it contains a large number of nodes, edges and loops. Our results reveal that sophisticated inference methods are essential for exploring the modeling capabilities of more complex graph structures. This fact highlights the need to consider both the inference method and the graph

[1] http://research.microsoft.com/downloads

[2] http://people.csail.mit.edu/dsontag/



(a) LogitBoost

(b) CRF-1D

(c) CRF-2D

(d) CRF-3D

(e) Legend for the class labels

Figure 6: Classification comparison for 180 blast-holes classified into three rock types. The CRF predictions presented are results of the best performing inference method (MPLP).

complexity jointly when addressing difficult problems.

The CRF-based approach presents a smoothing effect which correlates better with the known geology of the area. The results presented a marked increase, about 8%, in overall performance provided by the CRF approach over the base classifier, used as benchmark. In the particular data set tested, LBP inference had convergence issues and did not perform well. TRW-S was the most efficient method, fast and accurate for most of the graph configurations, but had difficulty with the three-dimensional graph structure. Finally, MPLP converged to the most probable distribution in all cases, although much slower. In particular for the 2D and 3D graphs, MPLP provided the best results with much lower standard deviations in all metrics. Nevertheless, in regards to computational time, when considering each algorithm independently using the chain CRF (1D) as baseline, both MPLP and TRW-S scaled similarly to the increase in complexity in the graph structure. It is worth noting that if MPLP were not used, the three-dimensional CRF would be considered a bad idea, since the other inference methods failed to improve classification performance.

Determining the detailed geology in terms of lithology, mineralogy and rock strength is a complex task and requires interpretation of the available data. The resulting classification was more consistent with the labeling scheme used for training, which considered each hole as belonging to one category. While this is evidently not the "true" down-hole geology, it is an approximation that seems to provide enough spatial resolution in this case—with steeply dipping rock layers and relatively shallow holes. The CRF was able to learn a trade-off between biasing the classification towards single-labeled holes and retaining the level of detail about rock types, especially in the boundaries between different zones.

Given the complex analysis required to provide ground-reference labels, possible future directions are investigating how to explore the large volume of unlabeled data and how to handle the uncertainty in the training labels. Another natural research direction from this paper is to study alternative methods to obtain the CRF graph structure. The problem of learning the structure of the graph based on real data and taking into account previous knowledge about the lithology distribution seems challenging.

Acknowledgments

This work has been supported by the Australian Centre for Field Robotics and the Rio Tinto Centre for Mine Automation.

References

[Besag, 1975] J. Besag. Statistical analysis of non-lattice data. *Statistician*, 24(3):179–195, 1975.

[Cho *et al.*, 2010] T. S. Cho, S. Avidan, and W. T. Freeman. The patch transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1489–1501, 2010.

[Dietterich *et al.*, 2004] T. G. Dietterich, A. Ashenfelter, and Y. Bulatov. Training conditional random fields via gradient tree boosting. In *International Conference on Machine Learning*, 2004.

[Friedman *et al.*, 2000] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28(2):337–407, 2000.

[Globerson and Jaakkola, 2008] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *Advances in Neural Information Processing Systems*, pages 553–560, 2008.

[Gutmann and Kersting, 2006] B. Gutmann and K. Kersting. TildeCRF: Conditional random fields for logical sequences. In *European Conference on Machine Learning*, pages 174–185, 2006.

Table 1: Performance comparison of the inference algorithms on different graph configurations

		LBP	TRW-S	MPLP
LB	Acc	82.26 (±3.2)	82.26 (±3.2)	82.26 (±3.2)
	Pr	81.71 (±6.5)	81.71 (±6.5)	81.71 (±6.5)
	Re	80.25 (±7.0)	80.25 (±7.0)	80.25 (±7.0)
	F	80.89 (±6.0)	80.89 (±6.0)	80.89 (±6.0)
	AUC	85.22 (±2.1)	85.22 (±2.1)	85.22 (±2.1)
1D	Acc	81.83 (±3.4)	87.30 (±2.7)	87.30 (±2.7)
	Pr	81.34 (±6.7)	89.28 (±6.9)	89.28 (±6.9)
	Re	79.67 (±7.8)	84.31 (±10.1)	84.31 (±10.1)
	F	80.41 (±6.4)	86.29 (±3.9)	86.29 (±3.9)
	AUC	84.79 (±2.6)	88.29 (±1.1)	88.29 (±1.1)
2D	Acc	82.06 (±3.3)	87.70 (±5.2)	88.46 (±1.4)
	Pr	81.61 (±6.3)	91.34 (±8.1)	88.59 (±1.4)
	Re	79.82 (±8.1)	83.71 (±22.0)	85.17 (±12.0)
	F	80.61 (±6.3)	85.86 (±12.2)	86.57 (±6.3)
	AUC	84.91 (±2.5)	87.88 (±8.9)	89.26 (±3.6)
3D	Acc	76.61 (±8.7)	79.32 (±7.3)	90.18 (±2.5)
	Pr	79.09 (±11.6)	78.27 (±11.6)	89.74 (±3.8)
	Re	68.71 (±40.4)	74.74 (±26.6)	87.47 (±11.6)
	F	68.75 (±29.5)	75.57 (±19.6)	88.43 (±7.5)
	AUC	76.92 (±15.7)	81.32 (±12.0)	90.98 (±4.8)

Acc = accuracy; Pr = precision; Re = recall;
F = F-score; AUC = area under the ROC curve;
Standard deviations are shown in brackets

[Hastie *et al.*, 2009] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, New York, NY, USA, 2nd edition, 2009.

[He and Zemel, 2008] X. He and R. Zemel. Learning hybrid models for image annotation with partially labeled data. In *Neural Information Processing Systems*, pages 625–632, 2008.

[Itakura *et al.*, 2004] K. Itakura, J. Doki, K. Sato, Y. Ichihara, and H. Matsumoto. Visualization of roof rock geostructure around tunnel by analysis of mechanical data from MWD. In *Asian Rock Mechanics Symposium*, volume 1, pages 179–184, 2004.

[Kadkhodaie-Ilkhchi *et al.*, 2010] A. Kadkhodaie-Ilkhchi, S. T. Monteiro, F. Ramos, and P. Hatherly. Rock recognition from mwd data: A comparative study of boosting, neural networks, and fuzzy logic. *IEEE Geoscience and Remote Sensing Letters*, 7(4):680–684, 2010.

[Koller and Friedman, 2009] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[Kolmogorov, 2006] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

[Lafferty *et al.*, 2001] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *International Conference on Machine Learning*, pages 282–289, 2001.

[Monteiro *et al.*, 2009] S. T. Monteiro, F. Ramos, and P. Hatherly. Conditional random fields for rock characterization using drill measurements. In *International Conference on Machine Learning and Applications*, pages 366–371, 2009.

[Nocedal and Wright, 2000] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, New York, NY, USA, 2nd edition, 2000.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[Scoble *et al.*, 1989] M. J. Scoble, J. Peck, and C. Hendricks. Correlation between rotary drill performance parameters and borehole geophysical logging. *Mining Science and Technology*, 8:301–312, 1989.

[Sokolova and Lapalme, 2009] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4):427–437, 2009.

[Sontag *et al.*, 2008] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss. Tightening LP relaxations for MAP using message passing. In *Uncertainty in Artificial Intelligence*, pages 503–510, 2008.

[Sutton and McCallum, 2007] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar, editors, *Introduction to Statistical Relational Learning*, chapter 4. MIT Press, 2007.

[Szeliski *et al.*, 2008] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[Teale, 1965] R. Teale. The concept of specific energy in rock drilling. *International Journal of Rock Mechanics and Mining Sciences*, 2:57–73, 1965.

[Utt, 1999] W. K. Utt. Neural network technology for strata strength characterization. In *International Joint Conference on Neural Networks*, volume 6, pages 3806–3809, 1999.

[Wainwright and Jordan, 2008] M. J. Wainwright and M. I. Jordan. *Foundations and Trends in Machine Learning*, volume 1, chapter Graphical Models, Exponential Families, and Variational Inference, pages 1–305. 2008.

[Wainwright *et al.*, 2005] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717, 2005.

[Werner, 2010] T. Werner. Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1474–1488, 2010.