10-21 MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Sparse Image Reconstruction by Two Phase RBM Learning: Application to Mine Planning

Yanyan Mu
Centre for Intelligent Machines
McGill University
Connell Engineering Building, Room 333
3480 University Street
Montreal, Quebec, Canada H3A 2A7
yanyanmu@cim.mcgill.ca

Frank Ferrie
Centre for Intelligent Machines
McGill University
Connell Engineering Building, Room 441
3480 University Street
Montreal, Quebec, Canada H3A 2A7
ferrie@cim.mcgill.ca

Roussos Dimitrakopoulos
Department of Mining and Materials Engineering
McGill University
Adams Building, Room 107
3450 University Street
Montreal, Quebec, Canada H3A 2A7
roussos.dimitrakopoulos@mcgill.ca

Abstract

A key problem in mine planning is estimating the locations of underground ore bodies from a set of sparse core samples that span the area to be excavated. Data from each sample location are interpreted by a geologist and rendered as an image depicting the local ore distribution. The goal is to reconstruct these sparse samples into a dense image that can correctly account for the underground structure. From a computer vision perspective, this has the form of a sparse data reconstruction problem, and is often tackled using a stochastic reconstruction approach. However in the present case the nature of the data is such that most conventional approaches fall short. In this paper we introduce a stochastic reconstruction method that uses a Restricted Boltzmann Machine (RBM) architecture to solve the problem in a novel way. Specifically, it incorporates a two-phase learning approach that i) uses dense sample information available from already excavated areas of the mine to build a general appearance model, and then ii) conditions this model to account for the data in the core sample images. Reconstruction is then accomplished by sampling the distribution implicit in the in the RBM after learning. Our results show that this approach offers significant improvements to conventional stochastic reconstruction algorithms as the RBM is better able to learn the distribution underlying the sample data.

1 Introduction

The problem of reconstructing a complete image from a set of partial observations is well-known in the computer vision literature [4, 6, 11]. Here we consider an analogous problem in a completely different context, often referred to as *mining simulation*, where the problem is to reconstruct a dense two-dimensional field from a set of sparse samples. The latter correspond to mineral core samples that are excavated from the ground through drilling, where the goal is to infer a more detailed representation of the mineral distribution (i.e. cores → pixels). A variation of Gaussian Processes, Kriging [9], has been widely used as a standard reconstruction technique in many mining simulation applications. However, due to the complicated variations in natural structures encountered in Geo-statistics, the Gaussian model is not compatable and usually generates poor reconstructions. Hence stochastic reconstruction with non-Gaussian distributions becomes necessary.

Conventional algorithms used in mining simulation operate in a way that is analogous to image reconstruction, first building a model to fit the prior distribution over the configuration space of training patterns. Then for given conditional data (i.e. novel), the model generates samples associated with the prior for simulation.

Over the past decade several approaches along these lines have been developed using pattern based algorithms, e.g., FILTERSIM [18, 17], SIMPAT[1, 2], and WaveSim[5]. However these algorithms are limited in their expressive power due to limitations of the model, which essentially boils down to manifold learning on patterns sampled from the training images. Reconstruction from novel data is then accomplished by generating random patch-wise paths through the image and populating each patch with a pattern drawn from the manifold associated with the closest novel input. As will be shown later in the experiments, this approach interpolates poorly when data are very sparse.

Another approch is characterized by statistically driven algorithms such as SneSim [12] and MRF-based Simlation [16]. As stochastic algorithms, a common characteristic is that in order to compute the posterior distribution, one has to sum over a high dimensional configuration space. This is often intractable, necessitating the use of techniques such as Markov Chain Monte Carlo (MCMC) methods. In computer vision, the Gibbs sampler was introduced to the field as a variation of MCMC to address this problem in the context of image restoration [8]. As a Bayesian approach, the prior distribution of the image is modeled by a Markov Random Field. To generate restorations of a degraded input image, the MRF model runs the Gibbs chain with simulated annealing to locate the energy minimum in the configuration space.

However, the limited connectivity of the MRF model often limits its expressive power. As a variation of the MRF, the Restricted Boltzmann Machine (RBM) is a stronger model with fully connected visible and hidden layers, where the parameters of these connections (weights) can be learned from a separate training set. The RBM and its variations have also been applied to image restoration tasks [13]. In this example, a two stage approach was used to i) learn a general appearance model (prior) from a set of dense samples and ii) adapt the general model according to a set of sparse samples (novel). By sampling the resulting model one implicitly solves the reconstruction problem. However, one can take this a step further. By stochastically sampling the model one can generate a distribution of images (this is referred to as a *simulation* in the mining simulation literatute) which one can think of in a way similar to particle filtering.

This notion of "simulation" is important as the intention in mine planning is to plan for contingencies. For example, in estimating the width of a mineral seam one would be more interested in determining the envelope containing all estimates rather than the maximum likelihood solution. For the remainder of the paper, the terms *reconstruction* and *simulation* are used interchangeably.

In this paper, a two phase RBM based simulation (RBMSim) is proposed, where RBM learns a full distribution over the pattern configuration space and can effectively sample the entire distribution with a Gibbs sampler. While the MAP solution is ambiguous, RBM coupled with simulated annealing can be used to explore multiple local minima and generate a number of plausible hypotheses within a prescribed confidence interval. In addition, a two phase learning approach is introduced to balance generalization and specificity. This also helps the algorithm cope with the complexity of the mining data.

2 RBM and Pattern based simulation

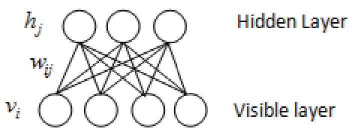

Figure 1. The parameters of a RBM consist of weight matrix w_{ij} associated with the connection between hidden units h_j and visible units v_i, also two bias weight vectors, b_i for the visible units and b_j for the hidden units respectively. In the context of image reconstruction, pixels correspond to visible units.

As shown in Figure 1, RBM is a two-layer undirected generative model, the stochastic binary units between two layers are fully connected. The joint configuration of units in both visible and hidden layers has an energy defined as (1):

$$E(v,h) = -\sum_{i,j} w_{i,j} v_i h_j - \sum_i b_i v_i - \sum_j b_j h_j, \quad (1)$$

where v_i and h_j are the states of visible and hidden units respectively, b_i and b_j are their bias, and w_{ij} are the weights between them.

By the energy configuration, the joint distribution of $p(v,h)$ is defined as (2):

$$p(v,h) = \frac{e^{-E(v,h)}}{Z}, Z = \sum_{v,h} e^{-E(v,h)}. \quad (2)$$

The marginal distribution of visible units is factorized as a product of exports (3):

$$p(v) = \frac{1}{Z} \prod_i e^{b_i v_i} \prod_j (1 + e^{\sum_i (b_j + w_{ij} v_i)}). \quad (3)$$

The activation of each visible unit given the hidden units and the activation of each hidden unit given the visible units are respectively defined as the following sigmoid functions (4, 5):

$$p(v_i = 1|h) = Sigmoid(b_i + \sum_j h_j w_{i,j}), \quad (4)$$

$$p(h_j = 1|v) = Sigmoid(b_j + \sum_i h_i w_{i,j}). \quad (5)$$

RBM learning finds the distribution of $p(v)$ that best represents the distribution underlying the training data by fitting the parameters w_{ij}, b_i, b_j. As with most undirected models, the RBM learning algorithm is based on variations of alternative Gibbs sampling. The log-likelihood gradient of RBM contains two terms (7, 8, 9), expectation w.r.t the data distribution (positive phase) and expectation w.r.t the model distribution (negative phase). The second term is intractable. Different learning methods use different algorithms to approximate this term.

$$\frac{\partial p(v)}{\partial w_{ij}} =< v_i h_j >_{p(h|v)} - < v_i h_j >_{p(v,h)}, \quad (7)$$

$$\frac{\partial p(v)}{\partial b_i} =< v_i >_{p(h|v)} - < v_i >_{p(v,h)}, \quad (8)$$

$$\frac{\partial p(v)}{\partial b_j} =< h_j >_{p(h|v)} - < h_j >_{p(v,h)} . \quad (9)$$

The most popular algorithm for learning is Contrastive Divergence (CD)[10]. To do the learning, a Gibbs chain is initialized from a sample of the training data, then alternated between hidden and visible units for several iterations. The last sample from the chain is used to approximate the intractable model expectation. In addition, Persistent Contrastive Divergence (PCD)[14] also uses Gibbs sampling to approximate the model expectation. But, PCD doesn't reset to the training data after weights update, the state of the chain *persists* from the previous iteration. By simply running a few iterations of the chain, these changes would be sufficient to track and sample from the *updated* model distribution. In addition, Fast Persistent Contrastive Divergence (FPCD)[15] aims to get a faster mixing of the Gibbs chain. It maintains a separate set of parameters that work with a higher learning rate for sampling only. The learning serves to push the chain out of the local mode.

Another noteworthy method to train RBM is Parallel Tempering (PT)[7] which uses Replica Monte Carlo. It runs several chains in parallel under different temperatures rather than a single Gibbs chain. As the temperature rises, the Gibbs distribution becomes smoother, facilitating mixing of the corresponding chain.

In brief, PT [7, 3] obtains a better approximation with less bias but heavier computational cost than CD [10, 14, 15]. In this paper, the RBM learning is implemented with PCD.

After learning, instances from $p(v)$ can be generated by running a Gibbs chain until it converges to an equilibrium distribution. This is accomplished using a block Gibbs sampling technique (Figure 2) which alternates activation of the hidden and active units such that when one is activated, the other is held fixed. For a sufficiently large t, the generated instance is guaranteed to be a sample drawn from $p(v)$.

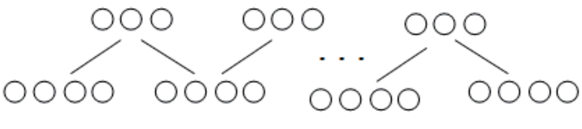

Figure 2. Block Gibbs Sampling.

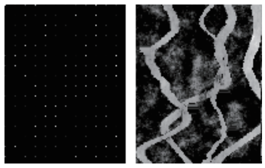

Figure 3. Task dependent data: (a) Trace of mineral seam determined by a geologist on the basis of core sample assay. (b) Synthetic image of ore body distribution rendered from core sample data.

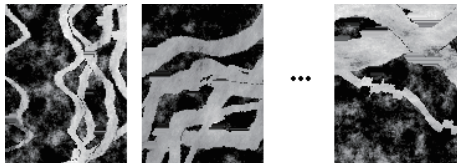

Figure 4. Task independent data: Global training set drawn from previously excavated areas.

3 Simulation by two phase RBM learning

Prior to excavation, a new area of the mine is sparsely sampled by drilling core samples and performing an assay on the core material. This analysis is based on expert knowledge and rendered in the form of an image. In fact, this sample-to-image encoding process is proprietary and performed for all measurements, both novel and prior. From a computer vision perspective, it provides normalization over the data which facilitates further analysis. Figure 3a shows the trace of a mineral seam determined by a geological expert from the analysis of sample data. On the right in Figure 3b is a rendering of the expected ore distribution produced by the encoding process. Note that is certain cases it is possible to generate alternate renderings for a particular set of data. In this paper there is only a single image per sample. We refer to these as the *task dependent data*.

Training an RBM solely with task dependent data will learn a distribution that best represents data in the vicinity of the core samples well, but generalizes poorly - particularly if samples are very sparse as is the case here. The solution adopted by all methods discussed here is to exploit the rich source of data available from prior excavated regions. In the case of a new installation, information is also available from archival databases drawn from other installations with similar characteristics. We refer to these as task independent data (Figure 4). Note also that because such information is gathered as part of the excavation process, measurements are available at full density.

The challenge is how to combine task dependent and task independent data, balancing specificity with the ability to generalize. In the case of systems such as WaveSim [5], the task dependent synthetic image and task independent data are all mixed together as the training set without distinguishing between local or global. The stochastic component is implemented by randomly selecting a pattern from within the closest manifold to the input sample. This produces good results if the core samples are reasonable dense, but interpolates poorly otherwise. The novelty of the RBM approach is in how it integrates these two sources of information to build a unified distribution that balances local and global.

This is implemented by a two phase learning process. RBM defines the distribution by three sets of parameters, the weights, w_{ij}, and two sets of biases b_i, b_j. The weight w_{ij} is the dominant parameter usually referred as the features. These reflect the natural characteristics of the ore distribution. The biases define how much the local training data favours the learned features. Hence the RBM is first trained with global training set, all parameters w_{ij}, b_i, b_j will be updated for each epoch. As these parameters converge, the second phase is initiated with the task independent data replaced by the task dependent data. However, in this phase only the biases of the hidden units, b_j, are updated at each epoch.

The process is summarised below:

- Building the Prior:
    - (1) Pattern Extraction
    - (2) RBM learning:
        - (2.1) First phase: Learn the weights and biases with task independent data.
        - (2.2) Second phase: Froze the weights and biases of visible units and learn the biases of hidden units with task dependent data.

- For simulation:
  (4) Define a patchwise random path that covers the sparse core sample image.
  (5) RBM sampling: Generate multiple simulations by sampling the RBM while clamping the task dependent data to visible units.

In the first phase, RBM learns generalized features; these features characterize a wide range of mining data. This phase will be trained once only, and as soon as training is complete the weights are frozen. In the second phase, the RBM adjusts the biases to “over-fit” the task dependent data to model the local characteristic of the ore distribution. This partitioning of RBM training appears to strike the appropriate balance between fit to data and reflecting the characteristics of the general population.

Reconstruction of the sparse data is effected by the simulation process in which a random path is generated that provides a complete cover of the image. This is done on a patch-wise basis using a grass fire algorithm to generate the path. Each image patch maps exactly onto the RBM depicted in Figure 1 with each visible unit, v_i, clamped to a specific pixel. The path is then traversed and block Gibbs sampling run to equilibrium at each patch along the way, yielding the simulation output. The exceptions are pixels corresponding to core samples (i.e. ground truth), which are initialized from the core sample data at the start of the process, but not updated.

Multiple simulation results are shown in Figure 5. The simulation results of WaveSim are also shown as a comparison in Figure 6.

We compare the RBMSim and wavesSim algorithms applied to data sampled at the Olympic Dam base metals deposit, South Australia. The two algorithms share the same training set which contains 30 images.

The advantages of the proposed RBM approach are clear upon examination of the experimental results. When input information is abundant, e.g. task dependent data derived from the core samples comprise approximately 5% or 10% of the grid, in Rows 3 and 4, the loci of the ore seams (brighter areas in the images) are similar. Both models tend to have a dominant local minimum on the energy surface. However, even in these cases the seams are better localized in the RBM output. From these observations we speculate that the RBM represents the prior distribution far more accurately than classical approaches typified by WaveSim.

A major difference in outputs occurs when the input drops to densities on the order of 1 or 2%. RBMSim is still able capture much of the structure visible in Rows 1 and 2, although there is some degradation in the fine structure. Each of these patterns corresponds to local minima of the energy surface. The WaveSim output, in contrast, loses much of the structural detail to the point where the output at 1 and 2% is uninformative. This implies that WaveSim cannot escape from the local minima resulting in simulations with limited stochastic variety.

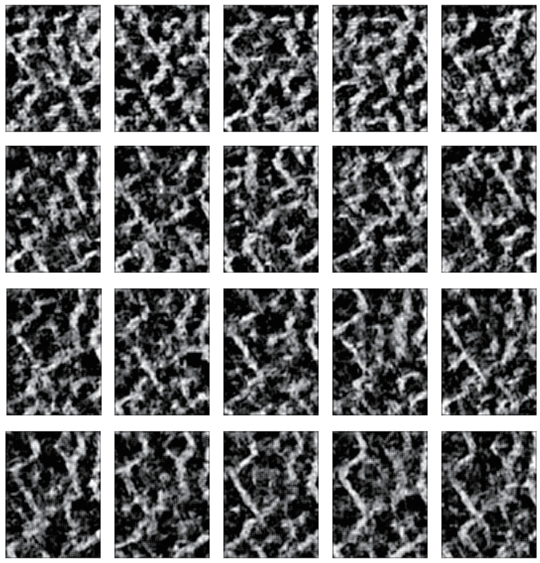

Figure 5. RBMSim simulation results: From top row to bottom are the simulations when core samples cover 1%,2%,5% and 10% of the area to be reconstructed. Each row contains 5 independent simulations.

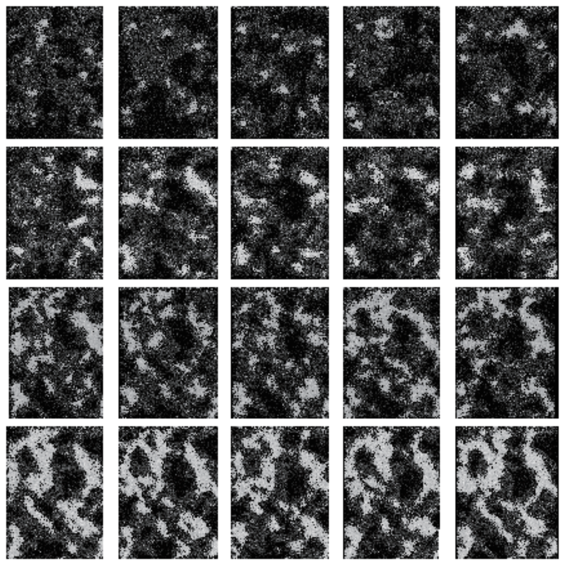

Figure 6. WaveSim simulation results: From top row to bottom are the simulations when core samples cover 1%,2%,5% and 10% of the area to be reconstructed. Each row contains 5 independent simulations.

4 Conclusion

This article focuses on the problem of reconstructing a dense estimate of an ore body distribution from a set of sparse core samples represented as images. It is approached as a stochastic image reconstruction problem and uses a Restrict Boltzmann Machine architecture to model the resulting distribution. What is novel about this work is that it uses an efficient, two phase learning approach to capture the general characteristics of the reconstructed data as well as the specifics conditioned by the sparse samples. Image reconstruction is effected by traversing a random path which covers the image while drawing from the learned distribution of the RBM.

In spite of the strength of the model, the main weakness is that the first learning phase is very time consuming. As the first learning phase is done, for each incoming task dependent data, the second phase learning and simulation are almost real time.

Another interesting aspect of this work concerns the idea of simulation. Since sampling of the RBM is a stochastic process, multiple solutions can be drawn that each fit the data to the same error (as reflected in the energy minima found using Gibbs sampling). This is important in the mining context as one desires and "envelope" about the true solution. In comparison to the current state of the art, the simulations presented here appear to be a significant improvement. It is also important to note that the accuracy of prediction has a significant impact on the cost of the mining operation, hence even an incremental improvement in algorithms can have a significant economic impact.

References

[1] G Burc Arpat and Jef Caers. A multiple-scale, pattern-based approach to sequential simulation. In *Geostatistics Banff 2004*, pages 255–264. Springer, 2005.

[2] G Burc Arpat and Jef Caers. Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203, 2007.

[3] Philemon B, Sander D, and Benjamin S. Training restricted boltzmann machines with multi-tempering: Harnessing parallelization. In AlessandroE.P. Villa, Wodzisaw Duch, Pter rdi, Francesco Masulli, and Gnther Palm, editors, *Artificial Neural Networks and Machine Learning ICANN 2012*, volume 7553 of *Lecture Notes in Computer Science*, pages 92–99. Springer Berlin Heidelberg, 2012.

[4] Marcelo Bertalmio, Luminita Vese, Guillermo Sapiro, and Stanley Osher. Simultaneous structure and texture image inpainting. *Image Processing, IEEE Transactions on*, 12(8):882–889, 2003.

[5] Snehamoy Chatterjee, Roussos Dimitrakopoulos, and Hussein Mustapha. Dimensional reduction of pattern-based simulation using wavelet analysis. *Mathematical Geosciences*, 44(3):343–374, 2012.

[6] Antonio Criminisi, Patrick Perez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *Image Processing, IEEE Transactions on*, 13(9):1200–1212, 2004.

[7] Guillaume Desjardins, Aaron C Courville, Yoshua Bengio, Pascal Vincent, and Olivier Delalleau. Tempered markov chain monte carlo for training of restricted boltzmann machines. In *International Conference on Artificial Intelligence and Statistics*, pages 145–152, 2010.

[8] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.

[9] Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Oxford university press, 1997.

[10] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

[11] Nikos Komodakis. Image completion using global optimization. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 442–452. IEEE, 2006.

[12] Sebastien Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002.

[13] Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey Hinton. Robust boltzmann machines for recognition and denoising. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2264–2271. IEEE, 2012.

[14] Tijmen Tieleman. Training restricted boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th international conference on Machine learning*, pages 1064–1071. ACM, 2008.

[15] Tijmen Tieleman and Geoffrey Hinton. Using fast weights to improve persistent contrastive divergence. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1033–1040. ACM, 2009.

[16] Haakon Tjelmeland and Jo Eidsvik. Directional metropolis: Hastings updates for posteriors with nonlinear likelihoods. In *Geostatistics Banff 2004*, pages 95–104. Springer, 2005.

[17] Jianbing Wu, Tuanfeng Zhang, and Andre Journel. Fast filtersim simulation with score-based distance. *Mathematical Geosciences*, 40(7):773–788, 2008.

[18] Tuanfeng Zhang, Paul Switzer, and Andre Journel. Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38(1):63–80, 2006.