Predicting ground state configuration of energy landscape using graph neural network

Seong Ho Pahng
Harvard University
spahng@g.harvard.edu

Michael P. Brenner
Harvard Universitys
brenner@seas.harvard.edul

Abstract

We describe a framework for finding ground state configurations of an ensemble of energy landscapes sharing distinct topological features. We construct a model energy landscape ensemble with Ising model Hamiltonian where the interaction matrices are drawn from protein contact map data. Our approach leverages graph neural network to learn a mapping from an interaction matrix J to a ground state configuration, yielding guesses for the set of most probable configurations. Given these guesses, we show that ground state configurations can be searched much faster than with vanilla simulated annealing. We also demonstrate that the search performance is sensitive to the complexity of the Hamiltonian.

1 Introduction

An energy landscape is a hypersurface which describes how the energy of a system changes as a function of system degrees of freedom [1]. Arguably one of the most important problems regarding an energy landscape is finding the ground state configuration. To this end, one can run simulated annealing to get Markov chains to anneal down to the global minimum as a simulation temperature approaches zero [2]. However, in cases where many interacting degrees of freedom result in a highly rugged energy landscapes, conventional sampling-based methods suffer from low probability of overcoming energy barriers and the chain may get stuck in local minima [3, 4].

In the physical sciences, we often have an *ensemble* of energy landscapes featuring similar topological patterns since there is a unique set of system degrees of freedom subjected to a single or handful of governing equations. One example is the energy landscapes of organic molecules built out of chemical building blocks where potential energies are obtained by solving Schrödinger equation. Under this ensemble setting, we investigate the possibility of using neural network to predict the ground state configurations of an energy landscape directly, circumventing the need to devise a system-specific sampling rule [5, 6].

To test the idea, we consider a model problem where we can define a natural ensemble. We construct Ising spin glasses [7], where the interaction matrix J is a *structured random* matrix. In this work, we instantiate this energy landscape ensemble using protein contact map as J matrices. Given the large database of natural proteins in Protein Data Bank [8], the protein contact map data gives a large set of J's upon which to define our ensemble. Recasting the global minimum prediction task as a node classification problem in graph theory, we employ a graph neural network (GNN) [9] to parametrize a mapping from a J to the corresponding ground state configuration. We then generate the set of most probable configurations from the GNN to predict low-lying configurations of the energy landscape. If these configurations miss the ground state configuration, we show that simulated annealing starting from a configuration in this set can search for the ground state configuration more efficiently. In addition, we probe how the GNN model performance is affected by the size and the topology of the energy landscape.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

2 Energy Landscape Ensemble

Model formulation Given an ensemble of similarly defined Hamiltonians, the resulting potential energy landscapes will feature similar patterns of undulation in high-dimension. To construct a model ensemble, we consider the following Ising model Hamiltonian

$$\mathcal{H}(\sigma) = -\frac{1}{2}\sum_{i,j}^{N} J_{ij}\sigma_i\sigma_j + h\sum_{i}^{N}\sigma_i, \quad h = \frac{\sum J_{ij}}{2N} \tag{1}$$

where both coupling and field terms depend on an interaction matrix J and $J_{ij}, \sigma_i \in \{0, 1\}$. The form of field is chosen to prevent all ground state configurations from trivially collapsing to a vector of all 1's. Under this formulation, the similarity of energy landscapes corresponds to the similarity of J matrices since J alone determines the Hamiltonian. We obtain an energy landscape ensemble by taking protein contact maps as our J ensemble. Proteins are characterized by distinct secondary structures with few additional non-local contacts, the pattern of which leads to a structured random connectivity in the J matrix. It should be noted that spin configurations here bear no relation to amino acid sequences; rather, the model represents a combinatorial optimization problem on a graph where a spin variable is turned on or off to minimize the potential energy given by Equation 1. We will thus use the term *node* for a spin variable σ from here on.

Dataset preparation We downloaded all *single-chain* protein structure files deposited in RCSB PDB [10] to ensure J's do not have a distinct block structure due to the presence of subunits and their features are solely governed by the pattern of intramolecular folding. We limit our data to proteins whose length ranges from 20 to 800 residues. We then generated a contact map following the standard procedure [11], and obtained 64563 contact maps to define a J ensemble. We ran a simulated annealing for each J starting from 100 random initial configurations, and selected an annealed configuration with the lowest energy as its purported ground state configuration, σ_{min}. Since the problem is NP-hard, we settle for this repeated annealing scheme and assume σ_{min} closely approximates the actual ground state configuration. We also counted the number of distinct annealed configurations from this experiment for future analysis. From all pairs of J and σ_{min}, 6400 pairs were randomly selected as a validation set, another 6400 as a test set, and the rest as a training set.

3 Computational method

Motivation for graph neural network Searching for ground state configurations in our model ensemble is a challenging task for two reasons. First, the connectivity set by J represents a quenched disorder in the Hamiltonian of Eq. 1, which, together with the field term offsetting the favorable energy contribution of turning on a node, gives rise to a frustrated energy landscape with many local minima. In addition, a configurational space to search over is very large, with the number of configurations growing as 2^N, where N ranges from 20 to 800. Due to these issues, we would like to forgo a stochastic search method and instead learn a mapping from J to a ground state configuration using GNN. Our idea is to generate an expressive node embedding from original node features and a graph structure, and use this embedding to classify globally whether each node shall be turned on or off. To allow for generalization of the mapping across J's, we chose a message passing framework [12] instead of a Laplacian-based convolution method which requires a constant graph structure.

Model architecture In this work, we applied Graph Attention Network [13] which utilize attention mechanism to assign different weights to different neighboring nodes. The network takes two input, an adjacency matrix J and node feature matrix, $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^M$ where M is initially 2 with node degree and field strength as two representative node features. At each network layer, we first transform these M features into M' hidden features via a weight matrix, $\mathbf{W} \in \mathbb{R}^{M' \times M}$. We then calculate an attention coefficient α_{ij}—which represents the influence of node j on node i—using a weight vector, $\vec{\mathbf{a}} \in \mathbb{R}^{2M'}$, as following

$$\alpha_{ij} = \frac{\exp\left(\sigma\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\sigma\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k]\right)\right)} \tag{2}$$

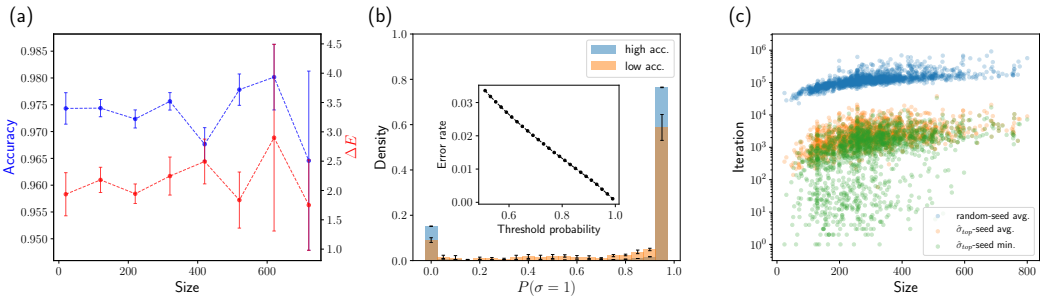

Figure 1: (a) Test set performance averaged over J's in a size window of 100. (b) Histogram of the classification probabilities. Inset shows the fraction of misclassified nodes among *confident* nodes. (c) Average number of sampling to reach σ_{min} in simulated annealing starting from a random configuration (blue) and $\hat{\sigma}_{top}$(orange). We also report the minimum number of samples (green).

where $\mathcal{N}_i$ denotes neighbors of node i, σ represents an activation function, and $\|$ a concatenation operation. The hidden feature of a node, $\vec{h}'$, is then computed by taking a weighted sum of hidden features of neighboring nodes. Employing multi-headed attention, we repeat $\vec{h}'$ calculation K times—each with different $\mathbf{W}$, $\vec{\mathbf{a}}$, and newly computed α—and concatenate them in an input and middle layers and averaged them in a final layer to retain M'=2, representing $P(\sigma_i$=0) and $P(\sigma_i$=1).

Training setup Unless otherwise mentioned, we followed the inductive learning setup of [13]. We trained the model to minimize the binary cross-entropy between the configuration probabilities and σ_{min} found via annealing, with a batch size of 16. After optimizing on the validation set, we found a five layer model with K=4 attention heads and M'=128 hidden features to perform best.

Enumerating top most probable configurations Since the original argmax point estimate scheme does not take a full advantage of the learned embedding, we generate a set of top most probable configurations from the configuration probability output of the GNN model. To obtain L such configurations, we pick top $\log_2 L$ nodes whose $P(\sigma$=1) are close to 0.5, and order all permuted configurations according to their corresponding sum combination of probabilities.

4 Results

As shown in Figure 1.(a), the fraction of correctly predicted nodes in $\hat{\sigma}$ decreases while the energy difference between a predicted configuration and σ_{min} increases as the size of J increases, the trend which is reasonable because large J's have a more complex pattern of connectivity than small J's with sparser connection. Notably, $\hat{\sigma}$ matched σ_{min} exactly in 1514 cases. The average accuracy and energy difference across the entire ensemble are 0.973 and 2.68 respectively, due to the size distribution skewed towards small J's. Figure 1.(b) shows the averaged histogram of $P(\sigma$=1) from high accuracy configurations in blue, and that of low accuracy configurations in orange. A striking feature is that most nodes in both cases were predicted with high certainty, as evinced by the peaks at both ends. In addition, the histogram of low accuracy configurations shows more nodes in the middle. To investigate whether model's prediction accuracy may be related to the node classification probability, $P(\sigma)$., we set a threshold probability, P_{thr} to select nodes whose $P(\sigma$=1) $\geq P_{thr}$ or $P(\sigma$=1) $< 1 - P_{thr}$, and calculated an error rate among these nodes as P_{thr} is varied. As shown in the inset of Figure 1.(b), the number of misclassified nodes among such nodes went down as we increase the threshold. This result in turn confirms that most misclassification indeed occur among *uncertain* nodes in the middle region of the histogram.

Given there are only a few uncertain nodes, we enumerated top 1000 most probable configurations for each J in the test set to account for $\log_2 1000 \approx 10$ uncertain nodes. We then calculated the energy of these configurations and picked the lowest energy configuration as an improved prediction, $\hat{\sigma}_{top}$, of the model. Through this configuration enumeration procedure, we found σ_{min} for additional 1658 J's. It also discovered configurations whose energy is lower than σ_{min} in 11J's, outperforming the round of dedicated simulated annealing.

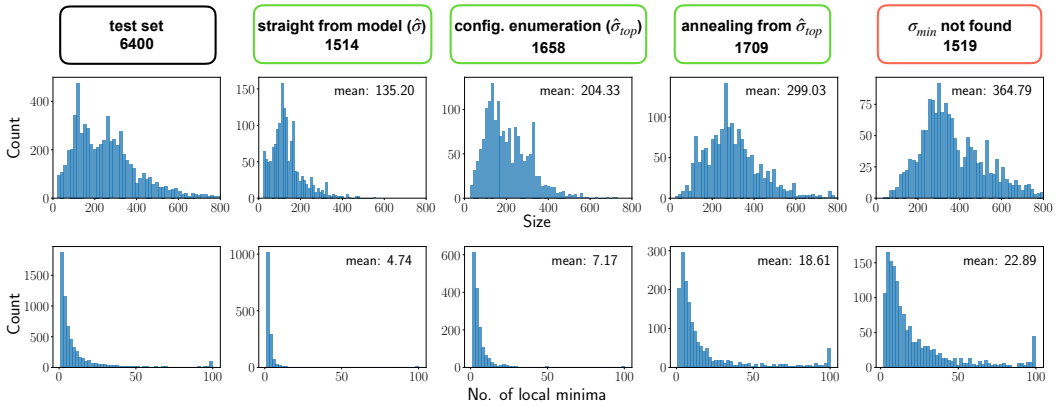

Figure 2: Histograms of size (1st row) and local minima (2nd row) for J's of the entire test set, and those belonging to the four performance categories.

For those J's where $\hat{\sigma}_{top}$ did not match σ_{min}, we ran simulated annealing with $\hat{\sigma}_{top}$ as a starting configuration and found σ_{min} for 1709 J's. As shown in Figure 1.(c), we observed about two orders of magnitude reduction in the number of sampling steps to reach σ_{min} for $\hat{\sigma}_{top}$-seeded simulated. The minimum number of steps found in $\sim 20\%$ cases was smaller than few hundreds as only one or two node were misclassified in $\hat{\sigma}_{top}$. It should be noted a simulated annealing can be launched from other configurations in the top most probable set to account for the possibility of $\hat{\sigma}_{top}$ falling in a basin that is far away from the one with σ_{min}.

To establish the relationship between the model's test set performance and the complexity of Ising energy landscape, we show in Figure 2 the size and local minima histograms of J's broken down into the following four categories: (1) where σ_{min} is found straight from the model via argmax operation, (2) where σ_{min} is found by taking the lowest energy configuration from the top 1000 most probable configurations (including the 11 special cases), (3) where σ_{min} is found by running simulated annealing from the configuration of (2), and (4) where σ_{min} is not found. We observe that the average size of J's and the average number of local minima of the associated Ising Hamiltonian both increase from (1) to (4), demonstrating the negative correlation between the accuracy of learned mapping and the complexity of the Ising energy landscape.

5 Discussion

Our work shows that it is indeed possible to use an ensemble of energy landscapes with known ground state configurations to train a neural network to deduce the ground state configuration of similar energy landscapes. On our model problem, we deterministically found the ground state configurations on 50% of the held out test set J's and stochastically on additional 26% through a graph neural network and seeded simulated annealing. Although this number may appear modest, we emphasize that all configurations predicted by the model were extremely low-lying configurations, often in the vicinity of the ground state configurations. Since the loss function does not include other local minima—or, for that matter, the energy function itself, we believe that such an informed prediction is possible only if the learned node feature embedding of GNN captures the complexity of the energy landscape. Accordingly, the GNN model's prediction suffers as the landscape complexity increases.

We also present the top configuration enumeration scheme to improve upon the model's initial prediction and cover a broad region at the bottom of an energy landscape, where every configuration is a viable candidate for simulated annealing seed. In fact, we expect many of ground state configurations in case (4) can be found via parallel tempering [14, 15] from these top configurations.

The contribution of our work is demonstrating that a neural network-based strategy can be used to find ground state configurations of an ensemble of energy landscapes. The next step is to apply the framework to a class of scientific problems where the discovery of global minima would have technological consequences.

Acknowledgments

We thank Yohai Bar-Sinai, Carl Goodrich, Mor Nitzan, and Jong Yeon Lee for helpful discussions. This work is supported by the Office of Naval Research through the grant N00014-17-1-3029.

References

[1] DJ Wales and TV Bogdan. Potential energy and free energy landscapes. *The journal of physical chemistry. B*, 110(42):20765–20776, 2006.

[2] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[3] Christos H Papadimitrou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[4] David P Landau and Kurt Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, 2014.

[5] Ulli Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361, 1989.

[6] Jian-Sheng Wang and Robert H Swendsen. Cluster monte carlo algorithms. *Physica A: Statistical Mechanics and its Applications*, 167(3):565–579, 1990.

[7] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[8] Helen Berman, John Westbrook, Zukang Feng, Gary Gilliland, TN Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[9] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. *arXiv preprint arXiv:1901.00596*, 2019.

[10] RCSB Protein Data Bank. URL `http://www.rcsb.org/`.

[11] Bohdan Monastyrskyy, Daniel d'Andrea, Krzysztof Fidelis, Anna Tramontano, and Andriy Kryshtafovych. Evaluation of residue–residue contact prediction in casp10. *Proteins: Structure, Function, and Bioinformatics*, 82:138–153, 2014.

[12] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[13] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[14] Robert H Swendsen and Jian-Sheng Wang. Replica monte carlo simulation of spin-glasses. *Physical review letters*, 57(21):2607, 1986.

[15] David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.