Deep Bayesian Optimization on Attributed Graphs

Jiaxu Cui[1,2], Bo Yang[1,2], Xia Hu[3]

[1] College of Computer Science and Technology, Jilin University, Changchun, China

[2] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, China

[3] Department of Computer Science and Engineering, Texas A&M University, College Station, United States

Motivation

1. Graphs are ubiquitous.
2. These real-world networks are often associated with a rich set of available attributes (The right figure is an illustration).
3. The cost of evaluating graphs is usually expensive.
4. It has been studied that the attributes on graphs are highly correlated to topological structures, and can benefit various network analysis tasks such as network embedding.

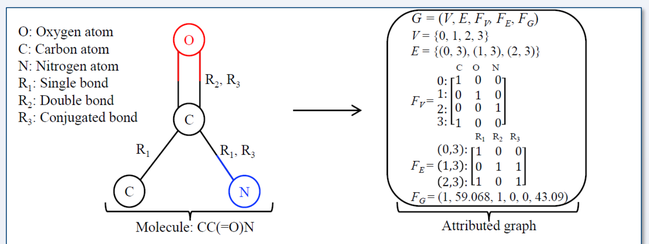


Figure 1: An illustration from a molecule sampling in Delaney data set (Delaney 2004) to an attributed graph (defined in Section 2). Node feature is atomic type, edge feature is chemical bond type, and global attributes contain minimum degree, molecular weight, # h-bond donors, # rings, # rotatable bonds, and polar surface area.

Questions To Be Explored

In this work, we propose to study **1) whether the attributes on graphs can benefit the task of graph structure optimization**; and **2) how to comprehensively explore available attributes to address this task more efficiently and effectively.**

Limitations of Current Methods and Difficulties

- Model-free methods usually require a large number of evaluations to maintain population diversity in finding an optimal solution.
- Existing vectorial Bayesian optimization (BO) methods cannot be directly applied to attributed graphs.
- Existing BO methods for, implicitly or explicitly, operating on graphs either are exclusively designed for the specific tasks such as neural architecture search, which are difficult to extend to other domains, or fail to consider all related attributes.
- The aforementioned model-based methods are mainly based on GPs (Gaussian processes), which is limited to its insufferable cubic scaling.

Proposed method: Deep Graph Bayesian Optimization (DGBO)

The iterative process of DGBO is as follows:

1. Train the deep surrogate model with training set;
2. Predict the properties of all candidates while capturing uncertainty via surrogate model;
3. Select a potential graph to evaluate next by maximizing the acquisition function;
4. Evaluate the black-box system to obtain the measure of the potential graph, until reaching a predefined termination condition.

The Proposed Deep Surrogate Model

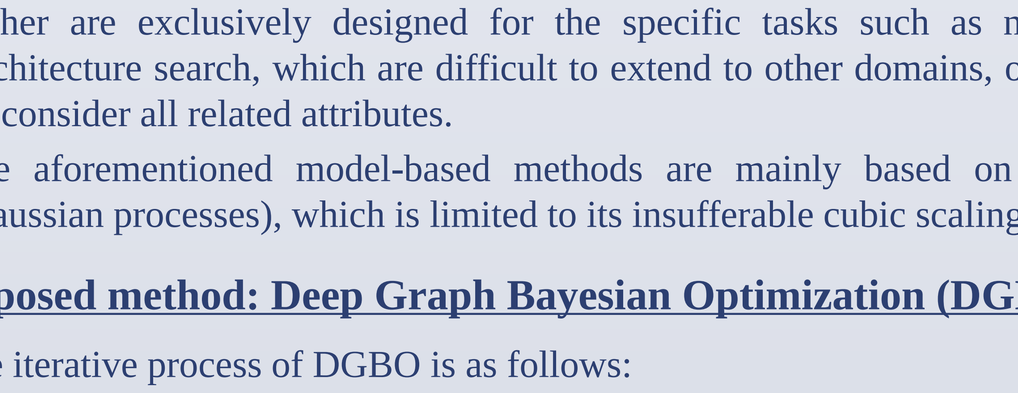


Figure 2: The overview architecture of the surrogate model in the DGBO. Its input is an attributed graph (e.g., a molecular graph) and output is a continuous measure (e.g., a desired property). F_v denotes the features of node v, F_e denotes the features of edge e and F_G denotes global attributes.

- Graph convolution (GC) layer:

$$H^{(l+1)} = \sigma(\sum_{r=1}^{D_E} \widetilde{D}_r^{-1/2} \tilde{A}_r \widetilde{D}_r^{-1/2} H^{(l)} W_r^{(l+1)})$$

- Pooling layer: $H^{(pool)} = \sigma(sum_{row}(softmax(H^{(l)} W^{(pool)})))$
- Prior layer: $H^{(con)} = Concat(H^{(pool)}, \lambda F_G)$
- Bayesian linear regressor (BLR) layer:

$$\mu(G^*) = \sigma_{noise}^{-2} \Phi(G^*)^T K^{-1} \phi(.) y_{1:N}$$
$$\sigma^2(G^*) = \Phi(G^*)^T K^{-1} \Phi(G^*) + \sigma_{noise}^2$$

Optimizing Surrogate Architecture by Transfer

In the real world, we often have very limited observations for the task at hand due to the high cost of function evaluation. To address this issue, in the paper we suggest employing the idea of transfer learning, i.e., to optimize surrogate architecture based on the available data from other sources.

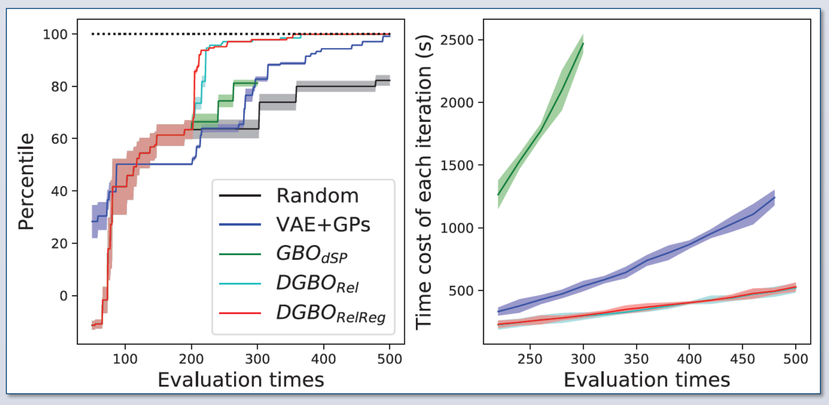
Parameters	Ranges	Optimal
# GC layers	{1, 2, 3, 4, 5}	5
# FC layers	{1, 2, 3, 4, 5}	5
# units of GC	[10, 100]	48
# units of pooling	[10, 100]	50
# units of FC	[10, 100]	45
$\sigma(.)$ of GC	$\{ReLU, tanH\}$	$tanH$
$\sigma(.)$ of pooling	$\{Identity, ReLU, tanH\}$	$Identity$
$\sigma(.)$ of FC	$\{Identity, ReLU, tanH\}$	$tanH$
Learning rate	[1e-4, 1e-1]	1e-4
Dropout	[0, 1]	0.0
Penalty coefficient	[1e-5, 1e-1]	1e-5

Table 1: The optimal surrogate architecture.

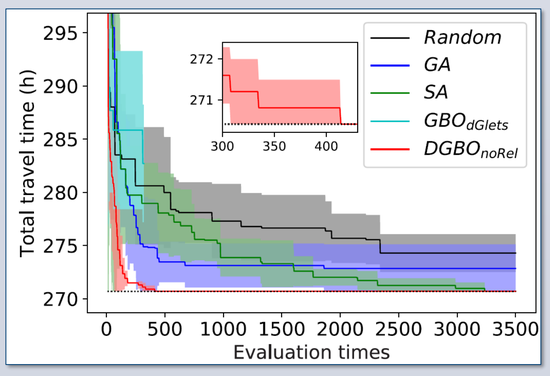
Experiments

We rigorously evaluate the proposed methods by answering three questions. 1) Can the available features from attributed graphs benefit optimization? 2) How effective and efficient are the proposed methods compared with the start-of-the-art on real-world problems? 3) Can they be applied to various domains? Specifically, we apply the proposed methods to an artificial non-linear function and two challenging real-world problems, including ***molecular discovery*** and ***urban road network design***.

- Results on molecular discovery :



- Results on urban road network design:



Conclusion

In this work, we propose a novel scalable global optimization method on attributed graphs, and apply it to solve various problems, including molecular discovery and urban road network design. The results show that the DGBO significantly outperforms the state-of-the-art methods in terms of both accuracy and scalability.

Contact: jxcui16@mails.jlu.edu.cn