論文の内容の要旨

論文題目　Modeling of Gene Regulatory Networks using Non Parametric Bayesian Models
（ノンパラメトリックベイズモデルを用いた遺伝子制御ネットワークのモデル化）

氏　　名　　パラフォクス　ノバック　レオン　フェリップ

In the last five decades, the human knowledge concerning the genetic network has grown exponentially, due in part to many new techniques used to obtain information from the genetic networks. Every day, new batches of data, describing the genome of bacteria, insects, animals and humans, are released to the public.
Clearly, our ability to process and interpret that amount of data has been surpassed by our ability to generate it. Many laboratories in the world are focusing their researches on processing and classifying all these genetic data. And among those, finding the regulations within genes, the gene regulatory network (GRN), has been of special interest.
Classical approaches have used known mathematical tools to attack a novel problem. Since those tools have already used and proven in other fields, is natural to explore their use in untested applications. Recently, the focus has been switching toward developing mathematical models that do not generalize to many models, but rather are very good at solving the problem presented at hand.
Particularly, the problem of Gene Network Inference requires a model that can be scalable, stochastic, and since the network connections are few, it also needs to be sparse. This means that it has to be able to model large matrices with few real values and many zeroes.
The algorithms used now were not designed to meet these requirements, simple ODEs have problems modeling stochastic systems, and it's hard to escalate them to higher dimensions due to the computational limitations. Also, these models do not have an inherent capability to model sparse data. ODEs model the network as a fully connected set, and is only later that any optimization problem fix the transcription rates of the connections.
Non Parametric Bayesian (NPB) methods, a tool first proposed in the 70's, provides a framework where we can model stochastic systems, regardless their size. NPB methods use infinite bounded priors to fit the data to a bounded model. They accomplish this by imposing infinite probabilistic process to known models like factor analysis or mixture of Gaussians. During the last decade, the use of NPB methods has been extending toward different areas of knowledge, like Natural Language Processing, Weather Prediction and Biology.

In this work, I propose the use of two NPB frameworks, the Dirichlet Process (DP) and the Indian Buffet Process (IBP), to create an ad-hoc model for GRN Inference. These processes both attack the problems of scalability, stochasticity and sparsity that the GRN problem has.
The DP clusters groups of co-regulated genes in a dynamic number of clusters, such that the model does not require any tuning if the test network changes. This lets the model have a dynamic clustering over different networks and samples.
The IBP, on the other hand, provides a prior capable of modeling a sparse network with infinite nodes. Using the Infinite Factor Analysis, the model can infer expressions closely regulated with regulatory genes. Furthermore, it also models binary and sparse networks as well.
Since both models are stochastic processes, the proposed model correctly addresses the desiderata when doing modeling of GRN.
In this work, I compare my model with conventional approaches, and provide an analysis of the model dynamics for different networks.
We use synthetic data to test the initial model, and then we use Benchmark Data from the DREAM challenge to test the model with the state of the art.
Finally, I present the results, and provide discussion on the future work and extension that could be done to further improve the model.