

Relatório Científico Parcial – Mestrado

Processo FAPESP 17/20575-9

Identificação de vias de sinalização celular baseada em
repositórios de cinética de reações bioquímicas

Beneficiário: Gustavo Estrela de Matos

Responsável: Marcelo da Silva Reis

Relatório referente aos trabalhos desenvolvidos entre 1 de janeiro e
10 de dezembro de 2018

Laboratório Especial de Toxinologia Aplicada, Instituto Butantan

São Paulo, 5 de Dezembro de 2018

Conteúdo

1	Resumo do Projeto Proposto	2
2	Introdução	2
3	Atividades Realizadas	4
3.1	Disciplinas cursadas	4
3.2	Estudo de literatura em seleção de modelos	4
3.3	Estudos de estimadores de verossimilhança marginal de um modelo	4
3.4	Implementação local do BIBm	4
3.5	Experimentos com o SigNetMS	4
4	Atividades futuras	4
4.1	Outras atividades acadêmicas	4
4.2	Participação em conferências	4
	Referências	4

1 Resumo do Projeto Proposto

A construção de modelos funcionais é uma técnica comum para se estudar vias de sinalização celular e, quando a via estudada é pouco conhecida, é possível que os modelos já propostos sejam incompletos, tornando necessário a sua modificação. Lulu Wu apresentou em 2015, em sua dissertação de mestrado, um método para modificar sistematicamente modelos funcionais, adicionando a estes interações extraídas de repositórios como KEGG. Entretanto, esta metodologia apresentou limitações: a primeira é a incompletude do banco de dados de interações criado, que extraia informações apenas do repositório KEGG; a segunda, a falta de informações sobre constantes de velocidade de interações, que podem ser extraídas de repositórios como BioNumbers; a terceira, a dinâmica do algoritmo de busca, incremental, que pode não achar o mínimo global; e a última, a penalização na complexidade dos modelos, que era feita de maneira aleatória. Propomos neste trabalho enfrentar as limitações encontradas pela metodologia de Lulu, criando um banco de dados de interações mais completo e também novas funções de custo que sejam capazes de penalizar modelos mais complexos (como critério de informação Akaike e *Bayesian inference-based modeling*); esta penalização deve induzir, em cadeias do espaço de busca, curvas em u no custo dos modelos, portanto também propomos a criação de novos algoritmos de busca que explorem essa característica da função de custo. Por fim, esperamos testar nossa metodologia na identificação de vias de sinalização celular da linhagem tumoral murina Y1.

2 Introdução

Vias de sinalização celular podem ser simuladas por modelos dinâmicos computacionais e, mais especificamente, modelos que descrevem a concentração de espécies químicas ao longo do tempo são chamados de modelos funcionais. Neste projeto, trabalhamos com modelos funcionais que descrevem as mudanças de concentrações de espécies químicas através de equações diferenciais ordinárias (EDOs). Estes modelos, quando não sofrem de sobreajuste

(*overfitting*), se tornam interessantes quando são capazes de reproduzir dados observados em experimentos biológicos, pois dão a estes modelos a qualidade preditiva.

O problema de criar modelos funcionais capazes de explicar resultados de experimentos biológicos com o mínimo de sobreajuste é chamado de *problema de identificação de vias de sinalização celular*. Podemos separar este problema em duas etapas principais. A primeira diz respeito a escolha da topologia da via de sinalização, o que é equivalente a escolher quais interações químicas são relevantes para o experimento em questão. A segunda etapa consiste em escolher valores para os parâmetros do sistema de EDOs do modelo funcional; estes valores são constantes de velocidade de interações e/ou concentrações iniciais de espécies químicas.

Em casos em que o experimento biológico ou a via de sinalização são muito estudadas, é possível que ambas as etapas descritas anteriormente possam ser resolvidas com pesquisas na literatura. Em casos que isto não é possível, torna-se uma solução recorrer a bancos de dados de biologia. Para a primeira etapa, podemos consultar bancos como o Kyoto Encyclopedia of Genes and Genomes (KEGG) [1], que contém mapas estáticos; e para a segunda, podemos consultar bancos como o BioModels [2] .

Entretanto, os mapas estáticos disponíveis nestes bancos podem ainda ser incompletos ou muito grandes para o experimento biológico em questão. Desta maneira, torna-se importante criar uma maneira sistemática de se modificar modelos funcionais a fim de fazê-los explicar o experimento biológico sem sobreajuste. Propomos fazer estas modificações tratando o problema de identificação de vias de sinalização como um problema de otimização combinatória, considerando como espaço de busca possíveis topologias para o modelo funcional e usando como função de custo alguma métrica que represente a qualidade deste modelo ao reproduzir o experimento biológico de interesse.

Considere S um conjunto de interações químicas. O conjunto de todas as possíveis escolhas de interações relevantes em S corresponde ao conjunto potência de S , $\mathcal{P}(S)$, que é o espaço de busca do problema de otimização que estamos interessados. Se chamamos a nossa métrica de qualidade de modelo funcional de c , transformamos nosso problema em

uma instância do problema de seleção de características. Assim, após definida a função de custo c e devidamente coletado e armazenado o conjunto S , propomos resolver instâncias do problema de identificação de vias de sinalização celular no arcabouço *featsel* [3].

3 Atividades Realizadas

3.1 Disciplinas cursadas

3.2 Estudo de literatura em seleção de modelos

3.3 Estudos de estimadores de verossimilhança marginal de um modelo

3.4 Implementação local do BIBm

3.5 Experimentos com o SigNetMS

4 Atividades futuras

4.1 Outras atividades acadêmicas

4.2 Participação em conferências

Referências

- [1] Minoru Kanehisa and Susumu Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.
- [2] Nicolas Le Novère, Benjamin Bornstein, Alexander Broicher, Melanie Courtot, Marco Donizelli, Harish Dharuri, Lu Li, Herbert Sauro, Maria Schilstra, Bruce Shapiro, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic

models of biochemical and cellular systems. *Nucleic Acids Research*, 34(suppl_1):D689–D691, 2006.

- [3] Marcelo S. Reis, Gustavo Estrela, Carlos Eduardo Ferreira, and Junior Barrera. feat-sel: A framework for benchmarking of feature selection algorithms and cost functions. *SoftwareX*, 6:193 – 197, 2017.