

Identification of cell signaling pathways based on biochemical reaction kinetics repositories

Gustavo Estrela de Matos

DISSERTATION PRESENTED
TO
INSTITUTE OF MATHEMATICS AND STATISTICS
OF THE
UNIVERSITY OF SÃO PAULO
FOR
THE DEGREE OF MASTER OF SCIENCE

Field of knowledge: Computer Science

Advisor: Dr. Marcelo da Silva Reis

Center of Toxins, Immune-Response and Cell Signaling (CeTICS)

Special Laboratory of Cell Cycle, Butantan Institute

During the development of this work the author received financial support of FAPESP.

São Paulo, February 7, 2019

Resumo

O problema U-curve é uma formulação de um problema de otimização que pode ser utilizado na etapa de seleção de características em Aprendizado de Máquina, com aplicações em desenho de modelos computacionais de sistemas biológicos. Não obstante, soluções propostas até o presente momento para atacar esse problema têm limitações do ponto de vista de consumo de tempo computacional e/ou de memória, o que implica na necessidade do desenvolvimento de novos algoritmos. Nesse sentido, em 2012 foi proposto o algoritmo Poset-Forest-Search (PFS), que organiza o espaço de busca em florestas de posets. Esse algoritmo foi implementado e testado, com resultados promissores; todavia, novos melhoramentos são necessários para que o PFS se torne uma alternativa competitiva para resolver o problema U-curve. Neste projeto propomos a construção de uma versão paralelizada e escalável do algoritmo PFS, utilizando diagramas de decisão binária reduzidos e ordenados. Além disso, propomos a criação de um novo algoritmo para o problema U-curve que se baseia na estratégia de divisão e conquista para particionar o espaço e achar a solução do problema de forma paralela. Os algoritmos desenvolvidos foram implementados e testados em instâncias artificiais e também em conjuntos de dados próprios para experimentos comparativos entre diferentes algoritmos de seleção de características.

Contents

| | | |
|---|------------------------|---|
| 1 | Introdução | 1 |
| 2 | Conceitos Fundamentais | 2 |
| 3 | Conclusão | 3 |

Chapter 1

Introdução

Chapter 2

Conceitos Fundamentais

Chapter 3

Conclusão

Bibliography

- [Lic13] M. Lichman. *UCI Machine Learning Repository*. 2013. URL: <http://archive.ics.uci.edu/ml>.
- [Rei12] M. S. Reis. “Minimization of decomposable in U-shaped curves functions defined on poset chains – algorithms and applications”. PhD thesis. University of Sao Paulo, 2012.
- [Wu15] Lulu Wu. “Um método para modificar vias de sinalização molecular por meio de análise de banco de dados de interatomos”. MA thesis. University of Sao Paulo, 2015.