System Perspective of Design Project

Motivation

To construct a Regulatory Network (RN), we have to struggle with preprocessing of multiple data sources, the number of which is growing; we may want to try different learning algorithms and experiment design and ranking algorithms; we may want to integrate newly available experiment results either from other studies or from our own studies into RN and we may want to establish a measurement to evaluate different RNs. All these complex steps fall under the umbrella of "the Construction of RNs". Thus we suggest Regulatory Network Construction System (RNCS), whose goal is to provide a systematic (as opposed to ad hoc) and computer-supported approach for discovering and understanding the regulation mechanisms between genes and proteins.

From a system point of view, the design project can serve as a perfect case study for the possibility of prototyping RNCS. While the main part of the design project is to build a specific RN and evaluate its effectiveness, the system part of it is to study the process, think of ways to automate different processes and gain experience to design a potential architecture for such a system.

System Description

In this section, we will describe the possible structure of RNCS. We have multiple sources of datasets (dataset 1 to dataset n), and annotated knowledge databases as inputs to the system. The machine should step through each of the following stages (as shown in Fig 1): pre-processing of dataset, inference of protein-protein, protein-DNA interaction using Machine Learning mechanisms, constructing RN, verifying the network using existing knowledge/hold-up data, design new experiments to verify/improve the network, feedback the result from the new experiments back to the system as source data and improve the network. Also, new results and knowledge from other independent studies and experiments can be fed and integrated into the system at real time.

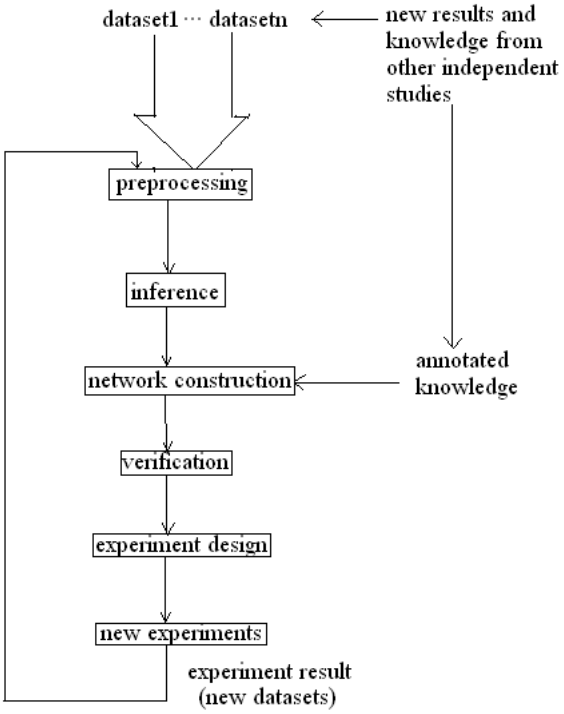

Fig 1: Regulatory Network Construction System (RNCS)

In addition, in order for such a system to run and reach a stable point automatically, we need to develop measures and matrices to evaluate the resulting network.

On the one hand, this system should be flexible enough to be able to integrate multiple resources and algorithms in each stage, be it pre-processing stage, inference stage or experiment designing stage. On the other hand, the network construction and improvement steps, shown in Fig 1, the close-loop feedback control system architecture and the evaluation mechanisms of the network should be common underlying stones.

Work Plan

The approach to weave the "System" perspective into this design project is by observing and experiencing the methods used in each construction stages, which can be systematical, manual or ad-hoc. Bearing in mind the ultimate goal as building RNCS and open issues, which are described in the next section, this case study can be broken down into several phases:

1. Study the process involved in constructing a RN during the project
2. Identify areas for automation and modularization by the end of the project
3. Describe the advantages of the proposed system architecture over current non-systematic approach

The extension upon successful completion of the design project is to document on the our experience with the whole process, literature review on other similar RN building studies and

research on the possibility of prototyping a RNCS

Open Issues

1. How to facilitate the integration of datasets (semantic queries (compare to semantic web), without knowing different components of the view/result is from different tables)
2. Is there any gaps that relational/traditional database can fill to benefit the data storage and access. Database is helpful in Selective Information Access (SIA) and cross-table information integration. Common commercial databases, also provide with a rich set of tools, which might be helpful
3. What degree of automation/modularization can be achieved (what processes)
4. Identify the common components and design an architecture for algorithm plug-in
5. Are there ways to automatically decide/provide hints for people to decide what algorithm to use in each stage
6. When should we claim the network is stable and run no more new experiment
7. How to compare and evaluate different networks (probability approach)
8. How to get out of the local maximum/optimum
9. How the result from new experiments or new datasets going to integrate to the existing network, thus rather than build a bunch of networks which needs biological expertise to identify them as useful/promising or not, we build and improve on a single network based on the cumulative experimental results or annotative knowledge
10. Is the integration of new results incremental or an aggregation (which needs either rerunning the system or partial reorganization which has the need of SIA)