Grand Valley State University

ScholarWorks@GVSU

2012

Translational Meta-Analysis Tool for Temporal Gene Expression Profiles

Guenter Tusch
Grand Valley State University, tuschg@gvsu.edu

Olvi Tole
Grand Valley State University

Follow this and additional works at: https://scholarworks.gvsu.edu/cispeerpubs

23rd International Conference of the European Federation for Medical Informatics
User Centred Networked Health Care - A. Moen et al. (Eds.)
MIE 2011 / CD / Demonstrations

Translational Meta-Analysis Tool for Temporal Gene Expression Profiles

Guenter TUSCH[a,1], Olvi TOLE[a]

[a] *Medical and Bioinformatics Program, School of Computing & Information Systems, Grand Valley State University, Allendale, MI, USA*

Abstract. Widespread use of microarray technology has led to highly complex datasets often addressing similar or related biological questions. In translational medicine research is often based on measurements that have been obtained at different points in time. However, the researcher looks as a progression over time. Our program (SPOT) helps the researcher find these patterns in large sets of microarray data. We created a software tool using open-source platforms and the Semantic Web tool Protégé-OWL.

Keywords. Research and education; enhancing biological, clinical and epidemiological research and trials; bioinformatics; translational research.

1. Methodology

Typically, for stimulus response studies a researcher obtains a fold change profile and tries to retrieve similar profiles in microarray databases or clinical databases (that more frequently include microarray data). However, this approach assumes that the pattern of time points is identical or at least very similar to the original experimental design. Some meta-analyses have used only the intersection of all time points or a selection with only the data available at that particular time. This is most likely only feasible if the studies are related. For our exploratory approach, it is more realistic to look, for instance, for a peak in the profile instead of correlating the entire profile. This can be accomplished by using knowledge-based temporal abstraction [1], dynamic time warping or spline interpolation approaches, where time-stamped data points are transformed into an interval-based representation. We implemented these ideas by creating a platform SPOT based on open-source software. It supports the R statistical package and knowledge representation standards (OWL, SWRL) using the Semantic Web tool Protégé-OWL connecting to the user through a web interface.

2. Design Principles

For the software implementation we chose the following design principles: 1. Use open source software as much as possible to avoid licensing and other problems so that our program will be available to everybody at no cost. 2. Follow Web 2.0 standards to ease interoperability of components. 3. A user interface that hides unnecessary technical

---

[1] Corresponding Author: tuschg@gvsu.edu.

detail from the user. 4. Web-based access for the most frequently used components to avoid tedious installation issues. 5. Re-use software through publicly available components to cut down on development time. 6. No use of non-local software components to avoid security leaks on the server side.
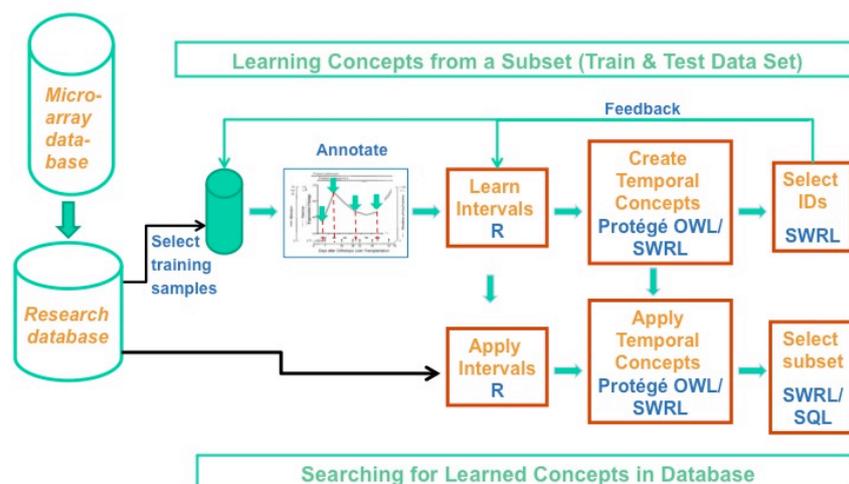

Figure 1. The SPOT Design.

3. System Functionality and Design

The user has to perform the following steps (figure 1): 1. Select datasets from the GEO database in the area of interest (figure 2). 2. Annotate time intervals in randomly selected raw data as "increasing", "decreasing", etc. (i.e., create training and test sets for algorithms). 3. Train algorithm (i.e. determine thresholds - multiple microarray platforms require multiple thresholds). 4. Chose from different algorithms for training patterns to recognize, for instance, a peak in fold changes of the temporal expression data. The program generates R code and OWL/SWRL rules. 5. Search for similar profiles in the database of interesting studies. For a more sophisticated analysis the data can be downloaded and analyzed locally. We developed tools to support this process using the Protégé-OWL ontology development toolkit. It supports knowledge representation standards (OWL, SWRL) using the Semantic Web tool Protégé-OWL (http://protege.stanford.edu/). The web interface is based on PHP and JavaScript running on an Apache server. A Java program interfaces R and Protégé. All statistical and GEO access functionality is implemented in R and Bioconductor using the Bioconductor package GEOmetadb to access GEO. The website is http://www.cis.gvsu.edu/~spot.

References

[1]   Tusch, G., Bretl, C., O'Connor, M., Das, A., SPOT--towards temporal data mining in medicine and bioinformatics, *AMIA Annu Symp Proc*. **2008**, 1157.