MATHEMATICAL MODELING AND BIG DATA ANALYTICS IN BIOMEDICINE

V. Brusic*

School of Science and Technology, Nazarbayev University, Astana, Kazakhstan; *vladimir.brusic@nu.edu.kz

The World's total data is doubling every two years. Data expansion includes growth in quantity, complexity, and types of data. The enormous rate of generation and on-line access to data is profoundly changing the way how business is conducted. Biomedical data include research and development data, clinical data, activity and cost data, patient behavior data, basic science data, and standards and ontologies, among others. Furthermore, Big Data approaches are increasingly needed for utilization of results from various Omics studies. Specific applications include predictive and content analytics that support drug discovery and optimization, the development of new diagnostic methods, and personalization of medicine. Biomedical data vary in granularity, quality, and complexity. There is a variety of sources and data formats - web pages, publications, technical reports, and databases. The challenge is to make the transition from data to actionable knowledge.

We propose the use of knowledge-based approaches whereby well-annotated data are combined with specialized analytical tools and integrated into analytical workflows. A set of well-defined workflow types with rich summarization and visualization capacity facilitates the transformation from data to critical information and knowledge that enable understanding, decision making, and selection of actions for solving various problems. The emerging Big Data requires dynamic integration of standardized data into knowledge bases and also make selected data sources accessible through integration with the analytical tools. The applications of big data analytics including proteogenomic applications for cancer profiling, proteomics applications for profiling T-cell immunome, and discovery of vaccine targets. In the initial step of Big Data analysis, data must be subject to quality control, error elimination, and filtering. The data is then analyzed and refined for different modeling purposes (eg. summarization, reporting, exploratory analysis, or discovery). The refined data is subsequently added to the data sources. Multi-level mathematical models are also integrated with the knowledgebase to perform simulations and predictions for predictive analytics within the knowledge-based systems (Figure 1). The Big Data is a resource that is increasingly important for the improvement of diagnosis, prognosis, and monitoring of the disease, facilitation of selection and optimization of therapies, and the improvement of human health.

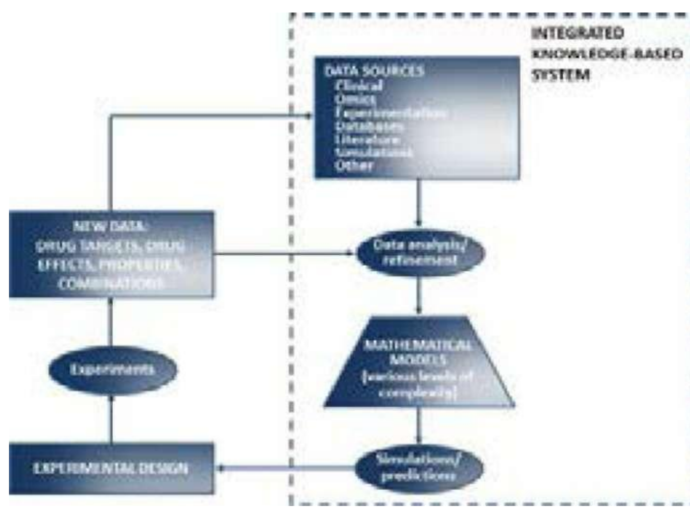

Figure 1. Integrated knowledge-based system

Multiple data sources (Fig. 1) provide inputs for selection of initial experiments that are performed to produce data needed for building of mathematical models. These models are in, turn, used to perform simulations that rapidly and effectively search many possibilities that suggest further experiments or provide possible descriptions of the observed systems. These data are then used for cyclical refinement of mathematical models and design of further experiments in a manner that saves time and money.