Expanding Data Mining Power with System Dynamics

E

Edilberto Casado
Gerens Escuela de Gestión y Economía, Peru

INTRODUCTION

Data Mining

Business intelligence (BI) is a key topic in business today, since it is focused on strategic decision making and on the search of value from business activities through empowering a "forward-thinking" view of the world. From this perspective, one of the most valuable concepts within BI is the "knowledge discovery in databases" or "data mining," defined as "the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques" (SPSS, 1997).

The usage of data mining as we currently know dates back to 1995 (Pyle, 2003). Since then, many applications were developed, and now it is a critical discipline to gain business insight. Table 1 shows a list of current data mining applications.

System Dynamics

System dynamics was created by Jay W. Forrester, Germeshausen Professor Emeritus of Massachusetts Institute of Technology (MIT), in 1956. It is defined by the System Dynamics Society (www.systemdynamics.org) as "a methodology for studying and managing complex feedback systems, such as one finds in business and other social systems."

System dynamics evolved from prior work in feedback control systems, and progressively its application was extended to fields other than engineering. Though its primary application is focused in the understanding of complex systems, it is also used as a predictive tool (Sterman 2000; An, Uhm, Kim & Kwak, 2002; Forrest, 1998). Table 2 shows a list of some fields where system dynamics has found applications.

LIMITATIONS OF DATA MINING FOR PREDICTIVE APPLICATIONS

Regarding data mining techniques and algorithms, Table 3 summarizes the most commonly accepted classification (Berson & Smith, 1997; Thearling, 2003; The Pilot Software's Data Intelligence Group, 1995).

As predictive tools, the data mining techniques listed in Table 3 have the following shortcomings:

1) *The statistical foundation of data mining:* Most of the current BI methods and tools, such as rule induction, decision trees, neural networks, and so forth, are extensively used to develop predictive models; and their conceptual foundation are a combination of mathematical, statistical, and artificial intelligence techniques. It is here where we find a source of limitation for a wider set of real-world applications, since statistics works with historical data and there is no full guarantee about predictions based on such data. A change in the characteristics

Table 1. A short list of data mining applications

- Credit scoring
- Customer behavior prediction
- Customer segmentation
- Database marketing
- Drug discovery
- Fraud prevention
- Non-compliant taxpayer detection
- Targeted marketing
- Web mining (analysis on Internet-generated data from a Web site)

Table 2. A short list of fields where system dynamics is applied

- Environmental sciences
- Social sciences
- Economics
- Software development
- Project management
- Enterprise management
- Supply chain management
- Health care
- Customer behavior

Table 3. Main categories of data mining techniques and algorithms

Decision Trees
Building of tree-shaped structures that represent sets of decisions.

Rule Induction
Extraction of useful if-then rules from data, based on statistical significance.

Nearest Neighbor
Classification of each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k > 1$).

Neural Networks
Statistical analysis tools that, through a "learning" process, build a model of the dependencies between the descriptive variables and the behavior to be explained.

Genetic Algorithms
Building of models using processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

of a market from a developing to a mature phase, for example, is enough to invalidate the use of statistical methods to forecast customer behavior (An et al., 2002).

2) *Explanation of results:* Another limitation of some data mining methods like neural networks, genetic algorithms, and so forth is their inability to provide an adequate explanation of their results, because they are not easily mapped into human terms (Moxon, 1996), or they are seen as black boxes and no explanation of the results is given, which inhibits confidence, acceptance, and application of results (Parallel Computer Centre at The Queen's University of Belfast, 1997).

3) *The thinking paradigm under the application of data mining methods:* As consequence of their statistical foundation, data mining methods provide information about trends and patterns in the context of a "straight-line thinking" paradigm (i.e., an outcome is expressed as a function of one or more independent variables). However, real world is a feedback system made of interacting elements in a closed-loop context, where an action leads to a result that affects current conditions, and the changed conditions become the basis for future action (Forrester, 1991). A growth strategy executed only relying in trends provided by data mining methods, for example, can lead to failures in achieving the desired goals, because the feedback effects in the business system were ignored (Avila, Mass & Turchan, 1995).

TOOLS OF SYSTEM DYNAMICS: BASIC CONCEPTS

There are two fundamental tools in system dynamics to represent the structure and behavior of a system: causal loop diagrams and stock-and-flow diagrams.

Causal Loop Diagrams

A causal loop diagram is a tool to represent the feedback structure of a system. It is a closed representation of the links between causes and effects involved in it, and also contains an identification of their most important feedback loops (Kirkwood, 1998; Sterman, 2000). Figure 1 shows an example of a causal loop diagram.

Stock-and-Flow Diagrams

A stock-and-flow diagram is another way to represent the feedback structure of a system. But, unlike a causal loop diagram, it can include more precise and detailed information about the nature of the system it represents. Figure 2 shows the following basic elements in a stock-and-flow diagram (Kirkwood, 1998; Sterman, 2000). While causal loop diagrams are more oriented to a qualitative description of a system, stock-and-flow diagrams allow relating mathematical functions to their elements and, therefore, they represent a quantitative description.

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/expanding-data-mining-power-system/14403

Related Content

Investigating Web 2.0 Application Impacts on Knowledge Workers' Decisions and Performance
Haya Ajjan, Richard Hartshorneand Scott Buechler (2012). *Information Resources Management Journal (pp. 65-83).*
www.irma-international.org/article/investigating-web-application-impacts-knowledge/70600

Organizational Aspects of Cyberloafing
Elisa Bortolaniand Giuseppe Favretto (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 2923-2928).*
www.irma-international.org/chapter/organizational-aspects-cyberloafing/14005

A Software Design Model for Integrating LMS and MOOCs
Talent T. Rugube, Colin Chibayaand Desmond Wesley Govender (2022). *Journal of Information Technology Research (pp. 1-14).*
www.irma-international.org/article/a-software-design-model-for-integrating-lms-and-moocs/299375

Modeling ERP Academic Deployment via Adaptive Structuration Theory
Harold W. Webband Cynthia LeRouge (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 2638-2645).*
www.irma-international.org/chapter/modeling-erp-academic-deployment-via/13959

Neural Data Mining System for Trust-Based Evaluation in Smart Organizations
T. T. Wong (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications (pp. 2950-2967).*
www.irma-international.org/chapter/neural-data-mining-system-trust/22856