Comp. by: RPrabhuGalleys0000836408 Date:26/5/08 Time:08:33:50 Stage:First Proof File Path://
ppdys1108/Womat3/Production/PRODENV/0000000005/0000008302/0000000016/0000836408.3D
Proof by: QC by:

W

What-If Analysis

Stefano Rizzi
DEIS, University of Bologna, Italy

Definition

In order to be able to evaluate beforehand the impact of a strategic or tactical move so as to plan optimal strategies to reach their goals, decision makers need reliable predictive systems. What-if analysis is a data-intensive simulation whose goal is to inspect the behavior of a complex system, such as the corporate business or a part of it, under some given hypotheses called scenarios. In particular, what-if analysis measures how changes in a set of independent variables impact a set of dependent variables with reference to a given simulation model; such a model is a simplified representation of the business, tuned according to the historical corporate data. In practice, formulating a scenario enables the building of a hypothetical world that the analyst can then query and navigate.

Historical Background

Though what-if analysis can be considered as a relatively recent discipline, its background is rooted at the confluence of different research areas, some of which date back to some decades ago.

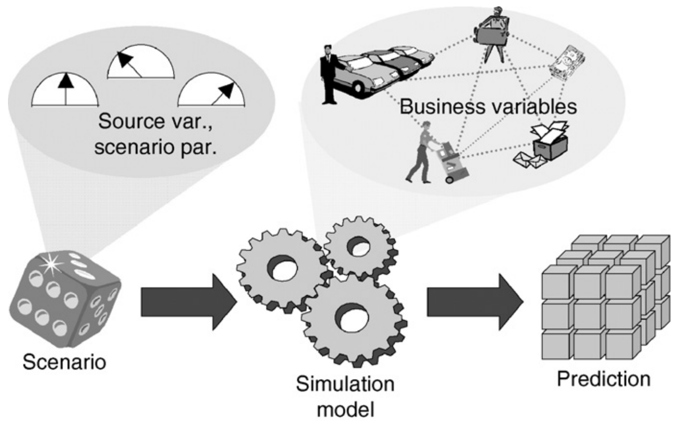
First of all, what-if analysis lends some of the techniques developed within the simulation community, to contextualize them for *business intelligence*. Simulations are used in a wide variety of practical contexts, including physics, chemistry, biology, engineering, economics, and psychology. A huge literature has been written in this field over the years, mainly regarding the design of simulation experiments and the validation of simulation models [1–3]. Au1

Another relevant field for what-if analysis is economics that provides the insights into business processes necessary to build and test simulation models. For instance, in [4] a set of alternative approaches to forecasting are surveyed, and useful guidelines for selecting the best ones according to the availability and reliability of knowledge are given.

Finally, what-if analysis heavily relies on database and *data warehouse* technology. Though data warehousing systems have been playing a leading role in supporting the decision process over the last decade, they are aimed at supporting analysis of past data ("what-was") rather than giving conditional anticipations of future trends ("what-if"). Nevertheless, the historical data used to reliably build what-if predictions are normally taken from the enterprise data warehouse. Besides, there is a tight relationship between what-if analysis and *multidimensional modeling* since input and output data for what-if analysis are typically stored within *cubes* [5]. In particular, in [6] the Sesame system for formulating and efficiently evaluating what-if queries on data warehouses is presented; here, scenarios are defined as ordered sets of hypothetical modifications on multidimensional data. Finally, there are relevant similarities between simulation modeling for what-if analysis and the modeling of *Extraction, Transformation and Loading* applications; in fact, both ETL and what-if analysis can both be seen as a combination of elementary processes each transforming an input data flow into an output.

Scientific Fundamentals

As sketched in Fig. 1, a what-if application is centered on a *simulation model*, that establishes a set of complex relationships between some *business variables* corresponding to significant entities in the business domain (e.g., products, branches, customers, costs, revenues, etc.). A simulation model supports one or more *scenarios*, each describing one or more alternative ways to construct a *prediction* of interest for the user. The prediction typically takes the form of a multidimensional *cube*, whose *dimensions* and *measures*-correspond to business variables, to be interactively explored by the user by means of any On-Line Analytical Processing (OLAP) front-end. A scenario is characterized by a subset of business variables, called *source variables*, and by a set of additional parameters,



What-If Analysis. Figure 1. Functional sketch for what-if analysis.

called *scenario parameters*, that the user has to value in order to execute the model and obtain the prediction. While business variables are related to the business domain, scenario parameters convey information technically related to the simulation, such as the type of regression adopted for forecasting and the number of past years to be considered for regression. Distinguishing source variables among business variables is important since it enables the user to understand which are the "levers" that she can independently adjust to drive the simulation. Each scenario may give rise to different simulations, one for each assignment of the source variables and of the scenario parameters.

A simple example of a what-if query in the marketing domain is: *How would my profits change if I run a* 3×2 (*pay 2 and take 3*) *promotion for one week on all audio products on sale?* Answering this query requires building a simulation model capable of expressing the complex relationships between the business variables that determine the impact of promotions on product sales, and to run it against the historical sale data in order to determine a reliable forecast for future sales. In particular, the source variables for this scenario are the type of promotion, its duration, and the product category it is applied to; possible scenario parameters could be the type of regression used for forecasting and the number of past years to be considered for regression. The specific simulation expressed by the what-if query reported in the text is determined by giving values "3×2," "one week" and "audio," respectively, to the three source variables. The prediction could be a *cube* with dimensions week and product and measures revenue, cost and profit.

Importantly, what-if analysis should not be confused with *sensitivity analysis*, aimed at evaluating how sensitive the behavior of the system is to a small change of one or more parameters. Besides, there is an important difference between what-if analysis and simple *forecasting*, widely used especially in the banking and insurance fields. In fact, while forecasting is normally carried out by extrapolating trends out of the historical series stored in information systems, what-if analysis requires simulating complex phenomena whose effects cannot be simply determined as a projection of past data.

On the other hand, applying forecasting techniques is often required during what-if analysis. In [4] the authors report a useful classification of forecasting methods into *judgmental*, such as those based on expert opinions and role-playing, and *statistical*, such as extrapolation methods, expert systems and rule-based forecasting. The applicability of these methods to different domains is discussed, and an algorithm for selecting the best method depending on the specific characteristics of the problem at hand is reported.

A separate mention is in order for *system dynamics* [7,8], an approach to modeling the behavior of nonlinear systems, in which cause-effect relationships between abstract events are captured as dependencies among numerical variables; in general, such dependencies can give rise to retroactive interaction cycles, i.e., feedback loops. From a mathematical standpoint, systems of differential equations are the proper tool for modeling such systems. In the general case, however, a solution cannot always be found analytically, so numerical techniques are often used to predict the behavior of the system. A system dynamics model consists of a set of variables linked together, classified as *stock* and *flow* variables; flow variables represent the rate at which the level of cumulation in stock variables changes. By running simulations on such a model, the user can understand how the system will evolve over time as a consequence of a hypothetical action she takes; she can also observe, at each time step, the values assumed by the model variables and (possibly) modify them. Thus, it appears that system dynamics can effectively support what-if applications in which the current state of any part of the system could influence its own future state through a closed chain of dependency links.

Designing a what-if application requires a methodological framework; the one presented in [9] relies on seven phases:

1. *Goal analysis*, aimed at determining which business phenomena are to be simulated, and how they will be characterized. The goals are expressed by (i) identifying the set of business variables the user wants to monitor and their granularity; and (ii) defining the relevant scenarios in terms of source variables the user wants to control.
2. *Business modeling*, which builds a simplified model of the application domain in order to help the designer to understand the business phenomenon as well as give her some preliminary indications about which aspects can be either neglected or simplified for simulation.
3. *Data source analysis*, aimed at understanding what information is available to drive the simulation and how it is structured.
4. *Multidimensional modeling*, which defines the multidimensional schema describing the prediction by taking into account the static part of the business model produced at phase 2 and respecting the requirements expressed at phase 1.
5. *Simulation modeling*, whose aim is to define, based on the business model, the simulation model allowing the prediction to be constructed, for each given scenario, from the source data available.
6. *Data design and implementation*, during which the multidimensional schema of the prediction and the simulation model are implemented on the chosen platform, to create a prototype for testing.
7. *Validation*, aimed at evaluating, together with the users, how faithful the simulation model is to the real business model and how reliable the prediction is. If the approximation introduced by the simulation model is considered to be unacceptable, phases 4–7 should be iterated to produce a new prototype.

The three modeling phases require a supporting formalism. Standard UML can be used for phase 2 (e.g., a use case diagram and a class diagram coupled with activity diagrams) and any formalism for conceptual modeling of multidimensional databases can be effectively adopted for phase 4. Finding a suitable formalism to give broad conceptual support to phase 5 is much harder, though some examples based on the use of colored Petri nets, event graphs and flow charts can be found in the simulation literature [10].

Key Applications

Among the killer applications for what-if analysis, it is worth mentioning profitability analysis in commerce, hazard analysis in finance, promotion analysis in marketing, and effectiveness analysis in production planning. Less traditional, yet interesting applications described in the literature are urban and regional planning supported by *spatial databases*, index selection in relational databases, and ETL maintenance in data warehousing systems.

Either spreadsheets or OLAP tools are often used to support what-if analysis. Spreadsheets offer an interactive and flexible environment for specifying scenarios, but lack seamless integration with the bulk of historical data. Conversely, OLAP tools lack the analytical capabilities of spreadsheets and are not optimized for scenario evaluation [6]. Recently, what-if analysis has been gaining wide attention from vendors of business intelligence tools. For instance, both SAP SEM (Strategic Enterprise Management) and SAS Forecast Server already enable users to make assumptions on the enterprise state or future behavior, as well as to analyze the effects of such assumptions by relying on a wide set of forecasting models. Also Microsoft Analysis Services provides some limited support for what-if analysis. This is now encouraging companies to integrate and finalize their business intelligence platforms by developing what-if applications for building reliable business predictions.

Future Directions

Surprisingly, though a few commercial tools are already capable of performing forecasting and what-if analysis, and some papers describe relevant applications in different fields, very few attempts have been made so far to address methodological and modeling issues in this field (e.g., [9]). On the other hand, facing a what-if project without the support of a design methodology is very time-consuming, and does not adequately protect the designer and his customers against the risk of failure. The main problem related to the design of what-if applications is to find an adequate formalism to conceptually express the simulation model, so that it can be discussed and agreed upon with the users. Unfortunately, no suggestion to this end is given in the literature, and commercial tools do not offer any general modeling support. Another relevant problem is to establish a general framework for estimating the loss of precision that is introduced when modeling low-level phenomena with higher-level dependencies. This could allow designers to assess the reliability of the prediction as a function of the quality of the historical data sources and of the precision of the simulation model.

Decision makers are used to navigating multidimensional data within OLAP sessions, that consist in the sequential application of simple and intuitive OLAP operators, each transforming a cube into another one. Consequently, it is natural for them to ask for extending this paradigm for accessing information also to what-if analysis. This would allow users to mix together navigation of historical data and simulation of future data into a single session of analysis. In the same direction, an approach has recently been proposed for integrating OLAP with data mining [11]. This raises an interesting research issue. In fact, OLAP should be extended with a set of new, well-formed operators specifically devised for what-if analysis. An example of such operator could be *apportion*, which disaggregates a quantitative information down a hierarchy according to some given criterion (*driver*); for instance, a transportation cost forecasted by branch and month could be apportioned by product type proportionally to the quantity shipped for each product type. In addition, efficient techniques for supporting the execution of such operators should be investigated.

Cross-references

► Business Intelligence, ► Data Warehousing Systems: Foundations and Architectures, ► Data Warehouse Applications, ► "Extraction, Transformation and Loading," ► On-Line Analytical Processing, ► Cube

Recommended Reading

1. Fossett C., Harrison D., and Weintrob H. An assessment procedure for simulation models: a case study. Oper. Res., 39(5): 710–723, 1991.
2. Kreutzer W. System Simulation – Programming Styles and Languages. Addison Wesley, Reading, MA, 1986.
3. Law A.M. and Kelton W.D. Simulation Modeling and Analysis. McGraw-Hill Higher Education, Boston, MA, 1999.
4. Armstrong S. and Brodie R. Forecasting for marketing. In G. Hooley and M. Hussey (eds.). Quantitative methods in marketing. International Thompson Business Press, London, 1999, pp. 92–119.
5. Koutsoukis N.S., Mitra G., and Lucas C. Adapting on-line analytical processing for decision modeling: the interaction of information and decision technologies. Decis. Support Syst., 26(1):1–30, 1999.
6. Balmin A., Papadimitriou T., and Papakonstantinou Y. Hypothetical Queries in an OLAP Environment. In Proceedigs of the VLDB Conference. Cairo, Egypt, 2000, pp. 220–231.
7. Coyle R.G. System Dynamics Modeling: A Practical Approach. Chapman and Hall, London, 1996.
8. Roberts E.B. Managerial applications of system dynamics. Pegasus Communications, 1999.
9. Golfarelli M., Rizzi S., and Proli A. Designing what-if analysis: towards a methodology. In Proceedings of the DOLAP. Arlington, VA, 2006, pp. 51–58.
10. Lee C., Huang H.C., Liu B., and Xu Z. Development of timed colour petri net simulation models for air cargo terminal operations. Comput. Ind. Eng., 51(1):102–110, 2006.
11. Chen B., Chen L., Lin Y., and Ramakrishnan, R. Prediction cubes. In Proceedings of the VLDB Conference. Trondheim, Norway, 2005, pp. 982–993.