Large-Scale LP in Business Analytics

William Chung
City University of Hong Kong, Hong Kong

INTRODUCTION

Although the definition of business analytics (BA) continues to change, BA generally involves big data and statistical and quantitative analyses to describe and investigate business performance. Results of the analyses are expected to provide new insights on future business planning by establishing several predictive models, which are also based on big data. Future business planning includes numerous operations and supply chain problems and solutions, such as demand forecasting, capacity planning, workforce planning, revenue management, inventory management, logistics analysis, and routing associated with prescription goals when the problem is optimization. Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, and Byers (2011) identified several merchandising problems, such as assortment optimization and pricing optimization. With the availability of big data, the aforementioned old-fashioned problems and solutions become new areas of research that might require the development of new and sophisticated BA approaches and methods. It is because big data generally refers to data that is so large and complex that they create significant challenges for traditional data management and analysis tools. For example, Weier (2007) reported that Wal-Mart handles 800 million transactions generated by its 30 million customers each day. Another chain store utilizes its data systems to handle a database table of more than four billion rows. Kiron and Shockley (2011) mentioned that the next big data measure after zettabyte (10^{21} byte) is yottabyte (10^{24} byte). Billions of years is required to download a yottabyte file at current high-speed broadband speeds. If Internet traffic continues to grow at current rates, we will likely approach the yottabyte milestone before the end of this century. IBM Big Data (2013) mentioned that 2.5 quintillion bytes of data are generated every day; 90% of the data in the world today has been generated in the last two years alone. This data can be considered as a kind of big data. Hence, the ability to store and retrieve big data is believed to be only a small piece of the competitive puzzle. Through several BA techniques, such as data mining and statistical prediction models, all successful retail Web sites can suggest additional purchases at checkout. Clearly, the size of the big data of companies is not the unique key factor to distinguish the winners. One of other key factors is the ability of companies to utilize optimization methods with big data to optimize resource utilization and make decisions, and one of the optimization methods is linear programming (LP). In this chapter, we discuss the application of large-scale LP from the BA perspective.

BACKGROUND

Large-Scale LP and Big Data

Linear programming (LP) has been applied in numerous fields after Danztig developed the simplex algorithm in 1947 to solve LP (Gass, 2011). Faster solution methods are desired when the size of LP problems becomes very large. Karmarkar (1984) invented a polynomial time algorithm for LP, and found that such algorithm is a class of interior point methods and may be faster than the simplex algorithm. The main difference between interior point methods and the simplex method

DOI: 10.4018/978-1-4666-5202-6.ch126

is the solution search path. The simplex method searches for the solution following the boundary of the feasible set, whereas the interior point method moves through the interior of the feasible region. We found that the interior point method is faster than the simplex method when the polyhedron of the feasible set is smooth and appears similar to a disco ball. The simplex method is faster than the interior point method when the polyhedron appears similar to a quartz crystal.

The term "large scale" is relative, and its meaning varies with time. For instance, the traveling salesman problem in the 1950s with approximately 50 cities (over one thousand variables) was considered large scale. The large-scale traveling salesman problem nowadays consists of more than 1000 cities (or over 10 million variables).

Bixby, Gregory, Lustig, Marsten, and Shanno (1992) utilized their experience to solve a large-scale, 12.75 million variable linear program by combining interior point method and simplex method. The researchers showed how the best features of the two methods could be combined to produce a very effective algorithm to solve a truly large-scale LP problem. However, the ever-increasing amount of readily available data in the modern, digital world requires solving optimization problems with millions of variables.

Sashihara (2012) presented a large-scale LP problem in a computerized decision-support system to maximize the contribution margin (revenue minus variable costs) of a flower-plant business problem in 2000. The actual set of equations in the computer program evaluated a solution matrix involving 120,000 rows and 420,000 columns. The result was that return on equity increased by 49%.

Within the context of big data, Tipi (2010) presented and discussed several large-scale (referred to as super-size in this presentation paper) industrial optimization problems. For instance, we may need to select among 200,000,000 maintenance routing options, schedule crews for 3400 daily flights in 40 countries, select 5 offers out of 1000 for each of 25 million customers, and so on. Super-size optimization problems involve up to tens of millions of constraints and up to hundreds of millions of decision variables. An extremely large number of columns (i.e., decision variables) would become much larger than the number of rows (i.e., constraints), and any computer memory could fit. Examples of such problems are crew pairing, sports scheduling, and vehicle routing.

In general, with the availability of big data, the size of large-scale LP is expected to become "very" large. Big data will provide more decision variables (columns) in LP. It should be noted that large-scale LPs could be classified into two groups: LP with and without special structure according to the solution methods. We first describe the larger-scale LP with special structure below. It is because there is a kind of scenario-based large-scale LPs with special structure whose problem sizes depend upon numbers of scenario generated from the availability of big data. That is, the scenario-based LP becomes a large-scale or very large-scale problem due to big data. More discussions will be given in the section of Main Focus.

Large-Scale LP with Special Structures

Problems with a moderate number of variables or constraints are considered large scale when the structures of the problems are difficult for existing solution methods. The basic idea of several solution methods for large-scale problems is to eliminate the problem that causes the difficulty. Every real-life problem is believed to have several exploitable structures. Known algorithms can be implemented to address difficult problems. Cutting planes method, Lagrangian relaxation method, Dantzig–Wolfe decomposition method, and Benders' decomposition method can be utilized to address difficult problems. Since the Dantzig-Wolfe decomposition method would be used to solve the scenario-based LP, we describe it below.

Related Content