European Journal of Operational Research 207 (2010) 1608–1619

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

A stochastic programming model for scheduling call centers with global Service Level Agreements

Thomas R. Robbins *, Terry P. Harrison

Department of Marketing and Supply Chain, College of Business, East Carolina University, Greenville, NC, United States
Department of Supply Chain and Information Systems, Smeal College of Business, Penn State University, University Park, PA, United States

ARTICLE INFO

Article history:
Received 9 June 2009
Accepted 11 June 2010
Available online 19 June 2010

Keywords:
Stochastic programming
Scheduling
OR in manpower planning
Call centers

ABSTRACT

We consider the issue of call center scheduling in an environment where arrivals rates are highly variable, aggregate volumes are uncertain, and the call center is subject to a global service level constraint. This paper is motivated by work with a provider of outsourced technical support services where call volumes exhibit significant variability and uncertainty. The outsourcing contract specifies a Service Level Agreement that must be satisfied over an extended period of a week or month. We formulate the problem as a mixed-integer stochastic program. Our model has two distinctive features. Firstly, we combine the server sizing and staff scheduling steps into a single optimization program. Secondly, we explicitly recognize the uncertainty in period-by-period arrival rates. We show that the stochastic formulation, in general, calculates a higher cost optimal schedule than a model which ignores variability, but that the expected cost of this schedule is lower. We conduct extensive experimentation to compare the solutions of the stochastic program with the deterministic programs, based on mean valued arrivals. We find that, in general, the stochastic model provides a significant reduction in the expected cost of operation. The stochastic model also allows the manager to make informed risk management decisions by evaluating the probability that the Service Level Agreement will be achieved.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

A call center is a facility designed to support the delivery of some interactive service via telephone communications; typically an office space with multiple workstations manned by agents who place and receive calls (Gans et al., 2003). Call centers are a large and growing component of the US and world economy, and are estimated to employ over 2 million call center agents (Aksin et al., 2007). Large scale call centers are technically and managerially sophisticated operations and have been the subject of substantial academic research. Call center applications include telemarketing, customer service, help desk support, and emergency dispatch.

Staffing is a critical issue in call center management, as direct labor costs often account for 60–80% of the total operating budget of a call center (Aksin et al., 2007). This paper addresses the scheduling problem in a call center with highly variable and uncertain arrival rates. The work is directly related to a research project with a provider of outsourced technical support delivered via globally distributed call centers. This operation involves providing help desk support to large corporate and government entities. While the scope of services varies from account to account, many accounts require 24×7 support and virtually all accounts are subject to some form of Service Level Agreement (SLA). There are multiple types of SLAs, but the most common specifies a minimum level of the Telephone Service Factor (TSF). A TSF SLA specifies the proportion of calls that must be answered within a specified time. For example, an 80/120 SLA specifies that 80% of calls must be answered within 120 seconds. A very important point is that the service level applies to an extended period, typically a week or month. Therefore, the desk is often staffed so that at some times the service level is underachieved, sometimes overachieved, and is on target for the entire month. The key challenge involved with staffing this call center is meeting a fixed SLA with a variable and uncertain arrival rate pattern.

Throughout this analysis, we will evaluate the models using three test problems based on specific outsourcing projects. Project J is a corporate help desk for a large industrial company averaging about 750 calls a day, where the volatility of call volume is relatively low. Project S is a help desk that provides support to workers in a large national retail chain. Call volume on this desk is about 2000 calls a day. Because this desk supports users in retail stores, as opposed to a corporate office, the daily seasonality of call volume is quite different.

* Corresponding author. Tel./fax: +1 252 328 6588.
E-mail address: robbinst@ecu.edu (T.R. Robbins).

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.06.013

This company is making major changes in its IT infrastructure and therefore call volume is very volatile and difficult to forecast. Project O is a help desk that provides support to corporate and retail site users of another retail chain. This is a smaller desk with about 500 calls a day, where call volume is fairly volatile and shocks are relatively common. We also examine various scheduling options. At one extreme, we only allow workers to be assigned to five 8-hour shifts per week. At the opposite extreme, we allow a wide range of part time schedules. We allow for a total of five different flexibility options (A-E), which are summarized in the Appendix in Tables A.1 and A.2.

2. Literature

There is a large body of literature addressing call center issues. Gans et al. (2003) provides a detailed and comprehensive review of the literature. Aksin et al. (2007) is another more recent review of the call center literature. Call center papers cover a wide range of topics and encompass a number of OR methodologies, including queuing theory, optimization, and simulation.

The primary problem we address in this paper is shift scheduling. The basic approach to this problem was first outlined in a paper by Dantzig (1954), which addressed scheduling toll booth operators. Dantzig formulated his model as a weighted set covering problem with known staffing requirements; the objective being to find the minimal cost covering from a set of available schedules. In the weighted set covering approach, the staffing levels in each time period are calculated exogenously and are defined as hard constraints that must be satisfied in any feasible schedule. Segal (1974) showed that without considering breaks the problem could be solved as a network flow problem in polynomial time. However, when breaks are scheduled explicitly the problem becomes NP Hard (Garey and Johnson, 1979). Due to the large number of potential schedules, especially when breaks are explicitly scheduled, much of the early research focused on solution algorithms.

Many early papers focused on heuristic algorithms. Henderson and Berry (1976) apply two types of heuristics. The first heuristic reduces the number of shift types, scheduling against only a reduced set of schedules referred to as the *working subset*. The second approximation is the scheduling algorithm, where the authors use three different scheduling heuristics. Another stream of research attacks the problem using an implicit scheduling approach. Implicit scheduling models use two sets of decision variables; one to assign breakless shifts, another to fit breaks. Implicit scheduling approaches are addressed in Bechtold and Jacobs (1990), Thompson (1995) and Aykin (1996). Several other papers address related problems (Brusco and Johns, 1996; Brusco and Jacobs, 1998, 2000). A succinct overview of a two-stage approach to scheduling in a call center environment is provided in Section 12.7 of Pinedo (2005).

Customer service is an important consideration in call centers, and many centers are subject to SLAs. Milner and Olsen (2008) examine contract structures in call centers with SLAs. Baron and Milner (2006) examine optimal staffing under various SLAs. These papers classify SLAs as Individual Based (IB), Period Based (PB), or Horizon Based (HB). IB-SLAs assess a financial penalty for every customer not served within the specified service level. The PB-SLA specifies penalties for each time period in which the service level target is not achieved. Periods are defined as intervals over which the arrival rate can be considered constant – typically 15 or 30 minute intervals. The HB-SLA specifies penalties for service level shortcomings over an extended period such as a week or month. In this paper we examine scenarios where a HB-SLA has been specified with the horizon specified as one week.

Most call center scheduling models in the literature implement a hard constraint for service level on a period-by-period basis – a PB-SLA. Scheduling for a PB-SLA is straightforward using the Stationary Independent Period by Period (SIPP) approach. The SIPP approach is described in detail in Green et al. (2001), but essentially the day is divided into short periods, typically 15 or 30 minutes. In each period, the arrival rate is assumed to be constant and performance is assumed to be independent of the performance in other periods. In each period, a queuing model, often the Erlang C model, is used to calculate the staffing level required to achieve the service level requirement. A set covering integer program is then used to schedule shifts. This two phased approach splits the task into a server sizing task, based on queuing models, and a staff scheduling task, based on discrete optimization.

A few models are formulated to solve a global service level requirement, i.e. an HB-SLA. It is our experience that outsourcing contracts often specify an HB-SLA, and all of the projects we examined were subject to this type of SLA. Koole and van der Sluis (2003) attempt to develop a staffing model that optimizes a global objective based on an HB-SLA. Their model uses a local search algorithm, and to ensure convergence to a global optimum they require agent schedules with no breaks, and assume no abandonment. Their model also assumes a time varying, but known, arrival rate. Cezik and L'Ecuyer (2007) solve a global service level problem using simulation and integer programming. They use simulation to estimate service level attainment and integer programming to generate the schedule. The IP model generates cuts via subgradient estimation calculated via simulation. The model solves the sample average problem and therefore ignores arrival rate uncertainty, but it does allow for multiple skills. This model is an extension of the model presented in Atlason et al. (2004). In a related paper Avramidis et al. (2007a) use a local search algorithm to solve the same problem. A related model is presented in Avramidis et al. (2007b). Fukunaga et al. (2002) describe a commercial scheduling application widely used for call center scheduling. Global service level targets are modeled as soft constraints while certain staffing restrictions are modeled as hard constraints. The algorithm uses an artificial intelligence based search heuristic. Atlason et al. (2008) develop an algorithm that combines server sizing and staff scheduling into a single optimization problem. This model focuses on the impact that staffing in one time period can have on performance in the subsequent period, a fact ignored in SIPP models. The algorithm utilizes discrete event simulation to calculate service levels under candidate staffing models, and a discrete cutting plane algorithm to search for improving solutions.

Each of these models either assumes that the per-period arrival rate is known or schedules against the expected arrival rate. The issue of arrival rate uncertainty has been addressed in several recent papers. Both major call center reviews (Gans et al., 2003; Aksin et al., 2007) have sections devoted to arrival rate uncertainty. Brown et al. (2005) perform a detailed empirical analysis of call center data. While they find that a time-inhomogeneous Poisson process fits their data, they also find that arrival rate is difficult to predict and suggest that the arrival rate should be modeled as a stochastic process. Many authors argue that call center arrivals follow a doubly stochastic process, a Poisson process where the arrival rate is itself a random variable (Chen and Henderson, 2001; Whitt, 2006; Aksin et al., 2007). Arrival rate uncertainty may exist for multiple reasons. Arrivals may exhibit randomness greater than that predicted by the Poisson process due to unobserved variables; the weather may have an impact on emergency calls (Chen and Henderson, 2001), the state of an organization's IT infrastructure may have an impact on support center calls (Robbins, 2007), and TV advertising may have an impact on inbound volume to a sales center (Andrews and Cunningham, 1995). Call volume is highly seasonal over the course of a day, week, month and year (Andrews and Cunningham, 1995; Gans et al., 2003; Robbins, 2007). Call center managers attempt to account for these factors when they develop forecasts, yet forecasts are subject

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات