A Multi-Objective Mathematical Model for Real-Time Employee Assignment in the Service Industry

Gercek Budak
Southern Illinois University Edwardsville

Xin Chen (✉️ xchen@siue.edu)
Southern Illinois University Edwardsville

Research Article

**Keywords:** Employee scheduling, mixed-integer programming, multi-objective optimization, service industry

**DOI:** https://doi.org/10.21203/rs.3.rs-842497/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
A Multi-Objective Mathematical Model for Real-Time Employee Assignment in the Service Industry

Gerçek Budak\textsuperscript{a}, Xin Chen\textsuperscript{b}

\textsuperscript{a}Visiting Scholar, Department of Industrial Engineering, School of Engineering, Southern Illinois University, Edwardsville, Illinois 62026-1805, USA; Email: gbudak@siue.edu

\textsuperscript{b}Corresponding author, Associate Professor, Department of Industrial Engineering, School of Engineering, Southern Illinois University, Edwardsville, Illinois 62026-1805, USA; Phone: 618-650-2853; Fax: 618-650-2555; Email: xchen@siue.edu

Disclaimer: This study does not directly involve any human or human data.
A Multi-Objective Mathematical Model for Real-Time Employee Assignment in the Service Industry

Abstract

The American economy has shifted toward services since the 1980s. The service industry is an important part of economy and is growing quickly in the last three decades. It is more human-capital intensive than the manufacturing sector and there is a shortage of highly-skilled workforce. One solution to this problem is to improve the efficiency through optimization. Because demand in the service industry changes constantly, it is a great challenge to determine the number of employees and their tasks to improve customer service while reducing cost. This article develops a multi-objective mixed-integer linear programming model to dynamically assign employees to different workstations in real time. A case study of the model is solved in less than one second and its pareto optimal solutions determine the number of employees who are assigned to each workstation and the expected customer service times. The mathematical model is robust and provides optimal employee assignment and service rates for workstations in many situations.

Keywords: Employee scheduling, mixed-integer programming, multi-objective optimization, service industry.

Declarations

Funding: Not applicable.

Conflicts of interest/Competing interests: No.

Availability of data and material: All data are included in the article.

Code availability: Available upon request.
1. Introduction

The American economy has shifted toward services since the 1980s (Hopp et al., 2007). For example, in the years of 1987, 1994, and 2001, 34.2%, 37.1%, and 42.7% of the national gross domestic product came from the service sector, respectively (Bureau of Economic Analysis, 2001). The service industry interacts with customers directly and extensively, and faces a great challenge when customer demand fluctuates in terms of the number of orders, products/services in an order, and the time orders are submitted. While many studies analyzed scheduling and throughput of mass production, only a few studies (e.g., Corominas et al., 2006; De Bruecker et al., 2018; Kwan et al., 1988; Vermuyten et al., 2018) investigated employee assignment and its impact on customer service rates in response to ever-changing customer demand.

With the advent of artificial intelligence and automation in the service industry, there is an increasing need to dynamically assign humans to workstations and provide customers with high-quality and timely services. This article addresses this challenge and develops a mathematical model whose solutions provide real-time employee assignment and ensure all workstations operate at the optimal pace to meet customer demand.

2. Background

Tasks in the service industry may be categorized into two different types: nondiscretionary and discretionary tasks (Hopp et al., 2007; Tan and Netessine, 2014). Nondiscretionary tasks are well defined and have relatively fixed completion times. Tasks in fast-food restaurants such as drive-through and take-out restaurants, post offices, and checkout lanes in supermarkets are mostly nondiscretionary tasks. Although these tasks have fixed completion times, customers may experience long waiting times before they receive service due to a large number of customers in
front of them in the queue. Discretionary tasks allow a certain degree of flexibility in task times and service providers adjust quality of task output and determine when tasks are completed. Tasks in sit-down restaurants, banks, call centers, and tasks performed by physicians, engineers, consultants, and other professionals are often discretionary tasks.

Tan and Netessine (2014) determined that servers in sit-down restaurants adjust their efforts between sales and service speed as workload changes. When workload is small (i.e., a server serves a small number of tables), the server expends more sales efforts with the increase in workload at a cost of slower service speed. Sales efforts include up-selling more expensive items (e.g., lobster tails and Kobe beef) and cross-selling additional items (drinks and desserts). Above a certain workload threshold, the server reduces sales efforts and increases service speed as the workload increases. There is a saturation point (i.e., optimal workload) that maximizes profit (higher sales and lower labor costs).

Mathematical models for manufacturing systems (e.g., Azizi et al., 2010) and the service industry (e.g., Corominas et al., 2006; De Bruecker et al., 2018; Vermuyten et al., 2018) were developed to assign tasks to multi-functional employees; each employee can perform a set of tasks. Heuristic solutions (e.g., Chua et al., 2017; Hur et al., 2004; Kwan et al., 1988; Vermuyten et al., 2018) were developed to schedule employees in the service industry. Mathematical models (e.g., integer programming; Choi et al., 2009) were developed to schedule both part-time and full-time employees in sit-down restaurants.
Most mathematical models use a static approach and schedule tasks and personnel at the beginning of a planning horizon (Mac-Vicara et al., 2017). These models assume that the supply and demand remain the same and employees scheduled to work are available for the planning period. In comparison, a small number of models (e.g., Hur et al., 2004; Mac-Vicara et al., 2017) dynamically assign tasks and personnel in response to stochastic demand and employee availability. These models revise schedules on a daily basis, but do not provide the granularity in minutes or hours to adjust tasks and reassign employees. In other words, these models are short-term scheduling tools.

Hur et al. (2004) conducted a detailed study of six McDonald’s franchise stores. Their study focused on determining the number of full-time and part-time employees at the beginning of each shift (work horizon) and used heuristics to find good solutions. A few other studies discussed how to assign workload based on customer orders in restaurants. These studies primarily focused on workload assignment in sit-down restaurants (e.g., Shimmura et al., 2011). This article aims to develop a methodology that assigns employees to workstations in real-time using historical demand data. A case study of the methodology is applied to a quick-service restaurant. The model may be expanded and applied to a variety of service companies where employee assignment and customer service rates are critical.

3. Problem Definition and Model Formulation

3.1. Problem Definition
In the service industry, demand is volatile and is met through interactions with customers. The schedule in the service industry determines the number and type (e.g., skills, full-time or part-time,
and other factors) of employees for a planning horizon, which varies from one shift to a few months. As demand (e.g., the number of customers who visit a store) increases and decreases over time, the need for service employees changes accordingly. Most companies in the service industry, however, tend to have a fixed schedule for employees over a long planning horizon. It is common to observe in a store that some employees are idle (e.g., employees at the deli department in a supermarket) while others are busy (e.g., employees working at checkout stations with long waiting lines).

Fast-food restaurants are among the most responsive service industry companies that dynamically schedule employees (Hur et al., 2004). They are able to determine the schedule (i.e., which employees will come to work at what time) right before the beginning of each shift, which lasts about four to five hours. Most tasks in a fast-food restaurant are nondiscretionary. A better schedule improves service quality by reducing customer waiting time and decreases cost by having fewer employees (Tan and Netessine, 2014). The case study in this article focuses on real-time scheduling for quick-service restaurants.

Many issues in the service industry may not be resolved through the scheduling of employees. For example, capacity is a major bottleneck for business. A gas station and/or electric car charging station can service a limited number of cars for fueling and/or charging at the same time. When more cars come to the station, they need to wait due to limited capacity. Other issues are caused by factors other than capacity or scheduling. For example, many patients experience long waiting times at physicians’ offices. There have been numerous studies (e.g., Marynissen and Demeulemeester, 2019) about scheduling in the healthcare industry. The capacity in a physicians’
office is well defined and the schedule may be optimized to minimize patients’ waiting time. The
tasks performed by physicians are highly discretionary. For example, a routine physical exam may
reveal major health issues that require immediate care and additional service time. The completion
time for discretionary tasks varies, which is one of the main reasons for long waiting times.

This article develops a mathematical model that enables real-time employee assignment in the
service industry. The optimal solution of the mathematical model indicates the number of
employees to be assigned to each workstation and the expected customer service rate for each
workstation. There are three main assumptions in this study: (a) task completion time is known
and is a parameter in the mathematical model. The completion time of a nondiscretionary task is
relatively certain with little variability and is obtained from historical data. The completion time
of a discretionary task has greater variability and may also be obtained from historical data. For
example, service times for patients scheduled for the next hour in a physician’s office are extracted
from historical data based on the type of appointments, the purpose of visits, gender, age, and other
factors; (b) quality of service does not improve with longer task completion time. For most service
companies studied in this research, longer task completion time is detrimental to service quality;
and (c) employees are multi-functional and may be assigned to any workstation. The case study
developed in this article is based on operations in quick-service restaurants, but the methodology
may be extended to provide real-time employee scheduling for many companies in both
manufacturing and service industries that have both discretionary and nondiscretionary tasks.

The multi-objective mathematical model developed in this research is a mixed-integer linear model
that aims to minimize (a) the maximum customer service time, including waiting time and task
completion time; (b) the total service time of all customers; (c) the number of workstations that contribute to customer service time exceeding the predefined service time; and (d) the total number of employees assigned to all workstations. This new mathematical model is different from other models developed earlier (e.g., Azizi et al., 2010; Corominas et al., 2006; De Bruecker et al., 2018; Vermuyten et al., 2018) and its main contributions lie in three areas. First, this model provides real-time decision support for employee assignment. It takes less than one second to compute an optimal solution to the model, which may be run on demand for any planning horizon. For example, if the service time for one customer becomes too long (i.e., the service time differs significantly from historical data), the model may be run to obtain a better optimal assignment immediately. Most previous models used heuristic approaches and did not provide optimal solutions (e.g., Chua et al., 2017; Hur et al., 2004; Vermuyten et al., 2018). In addition, previous models often required more computation time to calculate a solution (e.g., a few minutes in Hur et al., 2004), which prevented them from providing real-time decision support. These models were used for much longer planning horizons (e.g., days and weeks).

Secondly, the model developed in this article uses past demand data to compute an optimal schedule for the next planning horizon. For example, a manager may run the model at the beginning of every hour to determine the employee assignment for the next hour using the demand data from the previous hour or a time period in historical data that has similar demand pattern. For another example, when an increasing number of customers wait for long time to receive service, a manager may run the model and obtain an optimal solution immediately to reassign employees among workstations to reduce customer waiting times. The combination of the use of historical
operations data and short planning horizons helps provide better solutions to optimize service operations.

Thirdly, the model focuses on reassigning employees among workstations to reduce customer waiting times and operations costs. The model outcome may indicate that additional employees are needed to further improve service quality or some employees will be underload and may be assigned to taking a break, cleaning, accounting, inventory, or other tasks that do not service customers directly. This approach is vastly different from other models. Almost all previous research exclusively focused on determining how many employees should be assigned over a period of time such as a shift, day, week, or month. Once that decision was made, adjustment of workers among workstations in response to changing demand was not analyzed to improve efficiency and service quality.

3.2. Model Formulation
In the service industry, demand changes over time and managers need to make frequent decisions such as requesting employees pace up (increase service rate) or pace down (decrease service rate), reassigning employees from one workstation to another, and letting employees take breaks. When customer demand increases, decisions often involve pacing up and reassigning employees from certain workstations to the others. When customer demand decreases, decisions may include pacing down and assigning employees to tasks such as cleaning, accounting, and inventory. These decisions are for short periods of time (e.g., a few minutes to a few hours) and switch frequently in response to customer demand, employee availability, and other factors. Due to the relatively large number of tasks and employees, and the uncertainty and complexity (e.g., many types of
products/services) of customer orders, to find the best employee assignment and customer service rate is difficult. Managers make most decisions on the fly without real-time decision support.

The mathematical model for employee assignment and pace making is to solve an assignment problem with four objectives that minimize (a) the maximum customer service time; (b) the total service time of all customers; (c) the number of workstations that contribute to customer service time exceeding the predefined service time; and (d) the total number of employees assigned to all workstations. There are four sets (indices) in the model: \( i, j, k, \) and \( n \). They represent workstations, employees, products/services, and customer orders in a service system. They are defined as follows:

\( i \): Index of workstations in a service system, \( i \in I \), where \( I \) is a set of all workstations in the service system. There are total \(|I|\) workstations. \( i \geq 0 \). \( i \) is an integer and \( i = 1, 2, \ldots \).

\( j \): Number of employees assigned to a workstation in a service system. \( j \geq 0 \). \( j \) is an integer and \( j = 1, 2, \ldots \). There are total \( J \) employees in the service system.

\( k \): Index of products prepared and/or services provided in a service system, \( k \in K \), where \( K \) is a set of all products prepared and/or services provided in the service system. There are total \(|K|\) products. \( k \geq 0 \). \( k \) is an integer and \( k = 1, 2, \ldots \).

\( n \): Index of customers (orders) in time order who purchase products and/or acquire services in a service system, \( n \in N \), where \( N \) is a set of all customers in the service system. There are total \(|N|\) customers. \( n \geq 0 \). \( n \) is an integer and \( n = 1, 2, \ldots \).
There are six parameters in the model: $M, O_n, OT_{i,k}, T_{j,k}, TH_i$, and $Y_{k,n}$. Their values are known and obtained from historical demand data. These parameters are input to the mathematical model and are defined as follows:

$M$: A large integer number. $M \gg 0$.

$O_n$: The time at which the $n^{th}$ customer submits an order of products and/or services, or arrives at a service provider (the $n^{th}$ order). $O_n \geq 0$.

$OT_{i,k}$: A binary value indicating whether processing product/service $k$ requires workstation $i$. $OT_{i,k} = 1$ if workstation $i$ is used to process product/service $k$. $OT_{i,k} = 0$ if workstation $i$ is not used to process product/service $k$.

$T_{j,k}$: The time required to process product/service $k$ if there are $j$ number of employees. $T_{j,k} \geq 0$.

$TH_i$: The amount of time allowed for workstation $i$ in a service system to process products/services for any customer order. $TH_i$ is a target time specified by the service system; the time it takes to process any customer order at workstation $i$ is expected to be less than or equal to $TH_i$. $TH_i \geq 0$.

$Y_{k,n}$: The number of products/services $k$ in the $n^{th}$ order. $Y_{k,n}$ is an integer and $Y_{k,n} \geq 0$.

There are six variables in the model. Among them, two variables $u_i$ and $X_{i,j}$ are binary variables. The other four variables, $FT_{i,n}$, $MFT_n$, $MWT$, and $ST_{i,n}$ are nonnegative variables. These six variables are defined as follows:
\(u_i\): A binary variable indicating whether the time it takes for workstation \(i\) to process products/services ordered by any customer exceeds \(TH_i\). \(u_i = 1\) if the time exceeds \(TH_i\); otherwise \(u_i = 0\).

\(X_{i,j}\): A binary variable indicating whether \(j\) number of employees work at workstation \(i\). \(X_{i,j} = 1\) if \(j\) number of employees work at workstation \(i\); otherwise \(X_{i,j} = 0\).

\(FT_{i,n}\): The time at which workstation \(i\) completes processing products/services ordered by the \(n^{th}\) customer. \(FT_{i,n} \geq 0\).

\(MFT_n\): The time at which the \(n^{th}\) customer receives the order or leaves the service provider. \(MFT_n \geq 0\).

\(MWT\): Maximum customer service time (waiting time). \(MWT = \max_n (MFT_n - O_n)\).

\(ST_{i,n}\): The time at which workstation \(i\) begins processing products/services ordered by the \(n^{th}\) customer. \(ST_{i,n} \geq 0\).

The mathematical model has four objective functions (Eqs. (1)-(4)) and nine constraints (Eqs. (5)-(13)). Eq. (1) minimizes the maximum customer service time. Eq. (2) minimizes the total customer service time. When a queue in a service system is long or almost full, the majority of customers in the queue wait for almost the same amount of time to receive their orders (Shimmura et al., 2011). Eq. (1) is less attractive in this case since minimizing the maximum customer service time is equivalent to randomly selecting a customer and minimizing the customer’s service time. Eq. (2) is much more effective in alleviating this situation by minimizing the total service time of all customers. On the other hand, when a queue in a service system is short or almost empty, minimizing the maximum customer service time would reduce the waiting time of a few customers whereas other customers’ waiting time remains short. Eq. (1) becomes more effective in this case.
min \( MWT \) 

(1)

min \( \sum_n (MFT_n - O_n) \) 

(2)

min \( \sum_i u_i \) 

(3)

min \( \sum_i \sum_{j=1}^J (jX_{i,j}) \) 

(4)

Subject to

\( \sum_{j=1}^J X_{i,j} = 1, \ i = 1, 2, ... \) 

(5)

\( \sum_i \sum_{j=1}^J (jX_{i,j}) \leq J \) 

(6)

\( ST_{i,n} \geq O_n, \ i = 1, 2, ...; \ n = 1, 2, ... \) 

(7)

\( FT_{i,n} - ST_{i,n} - \sum_{j=1}^J (OT_{i,k,T_j,k,Y_{k,n}X_{i,j}}) \geq 0, \ i = 1, 2, ...; \ k = 1, 2, ...; \ n = 1, 2, ... \) 

(8)

\( MFT_n - FT_{i,n} \geq 0, \ i = 1, 2, ...; \ n = 1, 2, ... \) 

(9)

\( MWT - (MFT_n - O_n) \geq 0, \ n = 1, 2, ... \) 

(10)

\( FT_{i,n} - ST_{i,n+1} \leq 0, \ i = 1, 2, ...; \ n = 1, 2, ... \) 

(11)

\( FT_{i,n} - ST_{i,n} - Mu_i \leq TH_i, \ i = 1, 2, ...; \ n = 1, 2, ... \) 

(12)

\( M \sum_{j=1}^J (jX_{i,j}) - \sum_k \sum_n (OT_{i,k,Y_{k,n}}) \geq 0, \ i = 1, 2, ... \) 

(13)

\( u_i = 0 \ or \ 1, \ FT_{i,n} \geq 0, \ MFT_n \geq 0, \ MWT \geq 0, \ ST_{i,n} \geq 0, \ X_{i,j} = 0 \ or \ 1, \ i = 1, 2, ...; \ j = 1, 2, ..., J; \ n = 1, 2, ... \)

Eq. (3) minimizes the number of workstations whose customer service times exceed the predefined service time. In the service industry, many companies have a predefined service time and strive to make sure that customers receive their orders within the predefined service time (Hur et al., 2004).
Since most companies rely on multiple workstations to process customer orders, their predefined service time is broken into service times for workstations (the parameter $TH_i$ defined in this research). If workstation $i$ expends more time than $TH_i$ to process products/services in any customer order, $u_i$ in Eq. (3) becomes “1” for workstation $i$. The summation of all $u_i$’s indicates the number of workstations that exceed their respective predefined service times. A large value of Eq. (3) indicates certain workstations expand more time than expected to process products/services; employees from other workstations may be reassigned to these workstations to reduce service times.

Eq. (4) minimizes the total number of employees. While the service industry works hard to improve customer service, it is critical that companies reduce operations cost. A substantial amount of a service company’s cost is for human capital (e.g., cost of hiring and retaining employees). Research showed that profit and the number of employees have a nonlinear U-shaped relationship (Tan and Netessine, 2014). Too few or too many employees either decrease the revenue, or increase the cost, and lead to lower profit. Eq. (4) together with the other three objective functions (Eqs. (1)-(3)) provide a balance between customer service and profitability.

Eq. (5) is a constraint that determines the number of employees who work at workstation $i$. The binary variable $X_{i,j} = 1$ if $j$ number of employees work at workstation $i$; otherwise $X_{i,j} = 0$. Instead of using the binary variable $X_{i,j}$, an alternative approach is to use an integer variable (e.g., $X_i$) to determine the number of employees at workstation $i$. Both the binary and integer variables are discrete. The use of an integer variable reduces the total number of variables (i.e., total $|I| X_i$’s versus total $|I| \times J X_{i,j}$’s). Because the time it takes a workstation to process products/services
depends on the number of employees who work at the workstation, the mathematical model becomes a nonlinear model if an integer variable is used to determine the number of employees. For example, assuming it takes 100 seconds for workstation $i$ to process a product with one employee, and $X_i$ is an integer and represents the number of employees at workstation $i$. As $X_i$ may vary, it takes $\frac{100}{X_i}$ seconds to process the product. The term $\frac{100}{X_i}$ is nonlinear and makes the model a nonlinear model, which is difficult to solve except for special cases. In addition to $\frac{100}{X_i}$, other terms may be used to describe the relationship between service time and the number of employees and those terms are likely nonlinear if an integer variable is used. The use of a binary variable captures the nonlinear relationship between service time and the number of employees while ensure the mathematical model is linear and tractable.

Eq. (6) ensures that the total number of employees at all workstations does not exceed the maximum number of employees available. Eq. (7) indicates that a workstation begins processing products/services in a customer order after the order is submitted. Eq. (8) determines the time ($FT_{i,n}$) that workstation $i$ completes processing products/services in the $n^{th}$ customer order. In Eq. (8), the term $\sum_{j=1}^{I}(OT_{i,k}T_{j,k}Y_{k,n}X_{i,j})$ is the time it takes workstation $i$ to complete processing products/services in the $n^{th}$ customer order. $OT_{i,k}$ is a binary parameter indicating whether processing product/service $k$ requires workstation $i$. $T_{j,k}$ is a parameter indicating the time it takes $j$ number of employees to process one product/service $k$. $Y_{k,n}$ is a parameter indicating the number of products/services $k$ that are in the $n^{th}$ order. $X_{i,j}$ is a binary variable indicating the number of employees who work at workstation $i$. 


Eq. (9) determines the time \(MFT_n\) that the \(n^{th}\) customer order is completed (i.e., the time that all workstations have completed processing products/services in the \(n^{th}\) order). Eq. (10) determines the maximum service time \(MWT\) for customers, which is minimized in Eq. (1). Eq. (11) enforces the first-in-first-out (FIFO) principle. It dictates that any workstation may begin processing products/services in the \((n + 1)^{th}\) order after it completes processing products/services in the \(n^{th}\) order. The FIFO holds for most operations in the service industry. For example, in drive-through services (e.g., restaurants and banks), a customer who orders before another receives products/services before the other because customers form a line and those who arrive earlier are ahead of those who arrive later. There are exceptions in which later arrivals receive products/services earlier. For example, some drive-through restaurant services have temporary parking spots for customers. If a customer’s order is holding up the line, the customer might be asked to move to a temporary parking spot to wait for the order to be ready. Other customers come later and are behind this customer in the line may receive their orders before this customer.

Eq. (12) determines which workstations expend more time to process products/services than the predefined service times for the workstations. The total number of such workstations is minimized in Eq. (3). Eq. (13) ensures that if workstation \(i\) is required to process products/services in any customer order (i.e., \(\sum_k \sum_n (OT_{i,k}Y_{k,n}) > 0\)), a certain number of employees must be assigned to work at workstation \(i\) (i.e., \(\sum_{j=1}^J (jX_{i,j}) > 0\)). The mathematical model has a total of \(|I|(J + 1)\) binary variables and \(|N|(2|I| + 1) + 1\) nonnegative variables, four objective functions, and \(|I||N|(|K| + 4) + |I| + |N| + 1\) constraints.
As an example, in one hour of operations in a service company, six employees \((J = 6)\) work at five workstations \((|I| = 5)\) and process 30 customer orders \((|N| = 30)\). The company provides a total of 50 different products/services \((|K| = 50)\). The mathematical model uses the historical operations data for this hour to determine the optimal pace and employee assignment of each workstation in the next hour. Assuming that the same number of employees \((J = 6)\), workstations \((|I| = 5)\), and customer orders \((|N| = 30)\) will be available in the next hour, the model has 35 binary variables and 331 nonnegative variables, four objective functions, and 8,136 constraints. It is a multi-objective mixed integer linear programming model.

4. Case Study

The mathematical model is applied to a quick-service restaurant in Edwardsville, Illinois in the United States of America. In this restaurant, customers can either dine in or drive through. When customers dine in, they order and pay for the food, and pick up their food when it is ready. When customers drive through, they order the food, wait in the line, and pay for and pick up their food.

Both dine-in and drive-through services use the same workstations to process food items. The service speed of this restaurant is between those of a typical fast-food restaurant and traditional sit-down restaurant. In both dine-in and drive-through services, the service time is between the time a customer submits an order and the time the customer picks up the order.

This restaurant serves 50 different products \((|K| = 50)\), which are processed at seven workstations \((|I| = 7)\), including breakfast, cold beverage, dessert, frying, grilling, hot drinks, and salad. In the case study, the employee assignment and pace making model is applied at four representative times during a day that include breakfast time (9:00 AM), lunch time (12:00 PM), afternoon (4:00 PM),
and dinner time (7:00 PM). The model is solved using preemptive optimization in which the four objective functions (Eqs. (1)-(4)) are ranked by restaurant managers from the most to least important and solved sequentially. Table 1 below summarizes additional model parameters. In the case study, the demand data from last hour are used to compute the optimal assignment for the next hour. As discussed earlier, other historical data (e.g., the same time interval on previous days) may be used to compute the optimal assignment.

Table 1. Model Parameters for a Quick-Service Restaurant.

| Time of Decision | Number of Orders in One Hour (|N|) | Number of Employees in One Hour (J) |
|-----------------|--------------------------------|-----------------------------------|
| 9:00 AM         | 36                             | 8                                 |
| 12:00 PM        | 45                             | 10                                |
| 4:00 PM         | 10                             | 8                                 |
| 7:00 PM         | 60                             | 10                                |

Since this restaurant serves food faster than a sit-down restaurant but slower than a fast-food restaurant, it allows a maximum of 10 minutes for each workstation to complete processing food items in any customer order (\(TH_i = 10\)). Table 2 is the mapping between 50 food items the restaurant offers and seven workstations that process these items. Each food item is processed by one workstation and a workstation processes multiple food items. Please note that a food item may be processed by multiple workstations and this can be modeled using \(OT_{i,k}\). \(OT_{i,k} = 1\) if workstation \(i\) processes food item \(k\), which is marked by a “*” in Table 2. \(OT_{i,k} = 0\) if workstation \(i\) does not process food item \(k\). Workstations are numbered from 1 to 7 (\(i = 1, 2, ..., 7\)). \(i = 1\) is grilling, \(i = 2\) is frying, \(i = 3\) is cold beverage, \(i = 4\) is salad, \(i = 5\) is dessert, \(i = 6\) is breakfast, and \(i = 7\) is hot drinks. Food items are numbered from 1 to 50 (\(k = 1, 2, ..., 50\)).
Table 2. Food Items and Workstations ($OT_{i,k}$).

| $k$ | Food Items          | Workstations (i) | $k$ | Food Items          | Workstations (i) |
|-----|---------------------|------------------|-----|---------------------|------------------|
|     |                     | 1    2    3    4 | 6 7 |                      | 1    2    3    4 |
| 1   | Hamburger 1         | *     |     | 26 | Salad 2            | *     |
| 2   | Hamburger 2         | *     |     | 27 | Salad 3            | *     |
| 3   | Hamburger 3         | *     |     | 28 | Salad 4            | *     |
| 4   | Hamburger 4         | *     |     | 29 | Salad 5            | *     |
| 5   | Hamburger 5         | *     |     | 30 | Salad 6            | *     |
| 6   | Hamburger 6         | *     |     | 31 | Salad 7            | *     |
| 7   | Hamburger 7         | *     |     | 32 | Salad 8            | *     |
| 8   | Hamburger 8         | *     |     | 33 | Dessert 1          | *     |
| 9   | Hamburger 9         | *     |     | 34 | Dessert 2          | *     |
| 10  | Hamburger 10        | *     |     | 35 | Dessert 3          | *     |
| 11  | Fries 1            | *     |     | 36 | Dessert 4          | *     |
| 12  | Fries 2            | *     |     | 37 | Dessert 5          | *     |
| 13  | Fries 3            | *     |     | 38 | Breakfast 1        | *     |
| 14  | Fries 4            | *     |     | 39 | Breakfast 2        | *     |
| 15  | Chicken 1           | *     |     | 40 | Breakfast 3        | *     |
| 16  | Chicken 2           | *     |     | 41 | Breakfast 4        | *     |
| 17  | Chicken 3           | *     |     | 42 | Breakfast 5        | *     |
| 18  | Chicken 4           | *     |     | 43 | Hot Drink 1        | *     |
| 19  | Chicken 5           | *     |     | 44 | Hot Drink 2        | *     |
| 20  | Chicken 6           | *     |     | 45 | Hot Drink 3        | *     |
| 21  | Fish 1              | *     |     | 46 | Hot Drink 4        | *     |
| 22  | Fish 2              | *     |     | 47 | Hot Drink 5        | *     |
| 23  | Beverage 1          | *     |     | 48 | Hot Drink 6        | *     |
| 24  | Beverage 2          | *     |     | 49 | Hot Drink 7        | *     |
| 25  | Salad 1             | *     |     | 50 | Hot Drink 8        | *     |

The time it takes a workstation to process one food item depends on the number of employees who work at the workstation. Table 3 summarizes the time ($T_{j,k}$) of processing one food item $k$ with $j$ number of employees. A customer may order multiple of the same food item, which is the value of parameter $Y_{k,n}$. For example, suppose the 4th customer ($n = 4$) orders two Hamburger 1’s ($k = 1$), then $Y_{1,4} = 2$. Hamburger 1 is processed at workstation 1 (Table 2). Suppose there are three employees who work at workstation 1. Since $T_{3,1} = 1$ in Table 3, it takes workstation 1 two minutes to process two Hamburger 1’s ($T_{3,1}Y_{1,4} = 1 \times 2 = 2$). The values of the other two
parameters, the time the $n^{th}$ order is submitted, $O_n$, and the number of food item $k$’s in the $n^{th}$ order, $Y_{k,n}$, are included in Appendix A.

Table 3. Time (minutes; $T_{j,k}$) of Processing Food Items.

| Food Items ($k$) | Number of Employees ($j$) |
|------------------|---------------------------|
|                  | 0  | 1  | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   |
| 1                | Infinite | 5  | 2.5 | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 2                | Infinite | 6  | 3   | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 3                | Infinite | 6  | 3   | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 4                | Infinite | 6  | 3   | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 5                | Infinite | 6  | 3   | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 6                | Infinite | 7  | 3.5 | 2   | 2   | 2   | 2   | 2   | 2   | 2   |
| 7                | Infinite | 8  | 4   | 2   | 2   | 2   | 2   | 2   | 2   | 2   |
| 8                | Infinite | 8  | 4   | 2   | 2   | 2   | 2   | 2   | 2   | 2   |
| 9                | Infinite | 10 | 5   | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| 10               | Infinite | 10 | 5   | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| 11               | Infinite | 1.5| 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 12               | Infinite | 1.5| 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 13               | Infinite | 3  | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 14               | Infinite | 3  | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 15               | Infinite | 6  | 6   | 6   | 6   | 6   | 6   | 6   | 6   | 6   |
| 16               | Infinite | 6  | 6   | 6   | 6   | 6   | 6   | 6   | 6   | 6   |
| 17               | Infinite | 6  | 6   | 6   | 6   | 6   | 6   | 6   | 6   | 6   |
| 18               | Infinite | 6  | 6   | 6   | 6   | 6   | 6   | 6   | 6   | 6   |
| 19               | Infinite | 8  | 8   | 8   | 8   | 8   | 8   | 8   | 8   | 8   |
| 20               | Infinite | 8  | 8   | 8   | 8   | 8   | 8   | 8   | 8   | 8   |
| 21               | Infinite | 9  | 9   | 9   | 9   | 9   | 9   | 9   | 9   | 9   |
| 22               | Infinite | 9  | 9   | 9   | 9   | 9   | 9   | 9   | 9   | 9   |
| 23               | Infinite | 6  | 3   | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
| 24               | Infinite | 6  | 3   | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
| 25               | Infinite | 8  | 8   | 8   | 8   | 8   | 8   | 8   | 8   | 8   |
| 26               | Infinite | 10 | 5   | 5   | 5   | 5   | 5   | 5   | 5   | 5   |
| 27               | Infinite | 11 | 5.5 | 3   | 2   | 1   | 0.5 | 0.5 | 0.5 | 0.5 |
| 28               | Infinite | 12 | 6   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 29               | Infinite | 12 | 6   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 30               | Infinite | 12 | 6   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 31               | Infinite | 12 | 6   | 3   | 3   | 3   | 3   | 3   | 3   | 3   |
| 32               | Infinite | 12 | 12  | 12  | 12  | 12  | 12  | 12  | 12  | 12  |
| 33               | Infinite | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 34               | Infinite | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 35               | Infinite | 2  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 36               | Infinite | 2  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 37               | Infinite | 4  | 2   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
5. Model Results and Discussion

The mathematical model for the case study is solved using GAMS 24.0.2 on a desktop computer with 8.00 GB RAM and 64-bit Window 10 operating system. The solution time is less than one second. Table 4 below summarizes the model results for the case study. The priority of objective functions is determined by restaurant managers. In this restaurant and probably in many fast-food restaurants, managers strive to improve customer service, and minimizing maximum service time (Eq. (1); objective (1)) is their top priority. Managers rank minimizing the total number of employees (Eq. (4); objective (4)) as the least important objective. This might be due to the fact that fast-food restaurants do not have many high-paying jobs and reducing the number of employees does not have a significant impact on cost. The ranking of objectives (2) (Eq. (2)) and (3) (Eq. (3)) change depending on the time managers make decisions. Objective (2) emphasizes more on reducing total customer service time during busy times whereas objective (3) aims to smooth workload among workstations when certain customers experience long service times due to the food items they order. Managers prefer reducing total customer service time to balancing workload during busy times (12:00 PM and 7:00 PM in Table 4).
Table 4. Model Results for the Case Study.

| Priority of Objective Functions in Preemptive Optimization | Time of Decision |
|------------------------------------------------------------|------------------|
|                                                             | 9:00 AM | 12:00 PM | 4:00 PM | 7:00 PM |
| Objective Function (1) (Eq. (1); maximum service time in minutes) | (1)-(3)-(2)-(4) | 8.00 | 12.00 | 9.00 | 12.00 |
| Objective Function (2) (Eq. (2); total service time in minutes) | (1)-(2)-(3)-(4) | 251.00 | 293.00 | 71.00 | 326.50 |
| Total Number of Customers (Appendix A) | 36 | 45 | 10 | 60 |
| Mean Service Time (minutes) | 6.97 | 6.51 | 7.10 | 5.44 |
| Objective Function (3) (Eq. (3); number of workstations exceeding predefined service time) | 0 | 1 | 0 | 1 |
| Workstations Exceeding Maximum Service Time | None | 4 | None | 4 |
| Objective Function (4) (Eq. (4); total number of employees) | 5 | 10 | 6 | 10 |
| Total Number of Employees Available | 8 | 10 | 8 | 10 |
| Number of Employees Not Assigned to Workstations | 3 | 0 | 2 | 0 |
| Grilling (i = 1) | 0 | 2 | 2 | 2 |
| Frying (i = 2) | 0 | 1 | 1 | 1 |
| Cold Beverage (i = 3) | 1 | 2 | 1 | 2 |
| Salad (i = 4) | 0 | 1 | 0 | 2 |
| Dessert (i = 5) | 1 | 1 | 1 | 1 |
| Breakfast (i = 6) | 2 | 2 | 0 | 0 |
| Hot Drinks (i = 7) | 1 | 1 | 1 | 2 |

Table 4 shows that the mathematical model reduces the number of employees (three for 9:00 AM and two for 4:00 PM) during times when there are less customers. The model also optimally assigns employees to seven workstations. In some cases, however, managers do not have preference for the priority of different objectives, or they need to know the tradeoff between objectives. The mathematical may be revised to produce pareto optimal solutions (the efficient frontier) that comprise of a set of “optimal solutions” each of which has objective function values.
that are the same or better than that of any other solution. As an example, Table 5 shows the objective function values of all five pareto optimal solutions I-V computed for 7:00 PM.

Table 5. Pareto Optimal Solutions for Time of Decision at 7:00 PM.

| Pareto Optimal Solution | Eq. (1); Maximum Service Time in Minutes | Eq. (2); Total Service Time in Minutes | Eq. (3); Number of Workstations Exceeding Predefined Service Time | Eq. (4); Total Number of Employees |
|-------------------------|------------------------------------------|----------------------------------------|---------------------------------------------------------------|----------------------------------|
| I                       | 12.00                                    | 326.50                                 | 1                                                             | 10                               |
| II                      | 12.00                                    | 335.50                                 | 1                                                             | 9                                |
| III                     | 12.00                                    | 373.50                                 | 1                                                             | 8                                |
| IV                      | 12.00                                    | 415.50                                 | 1                                                             | 7                                |
| V                       | 12.00                                    | 455.00                                 | 1                                                             | 6                                |

In Table 5, the maximum service time (12.00 minutes) and the number of workstations exceeding predefined service time (1) remain the same. There is a tradeoff between the total service time and the number of employees. The implication of this tradeoff is that more employees should be assigned to workstations to improve service quality (decrease total service time). Since there are 60 customers who arrive between 6:00 PM – 7:00 PM (Appendix A), the mean service times for each customer from top to bottom in Table 5 are 5.44, 5.59, 6.23, 6.93, and 7.58 minutes, respectively. The corresponding numbers of employees are 10, 9, 8, 7, and 6, respectively. In other words, having one less employee increases the mean service time. The marginal impact on service quality is the least when the number of employees decreases from 10 to nine. If a manager would like to take both service quality and cost into consideration, having nine employees who work at seven workstations is the best choice.

6. Conclusions

The service industry is an important part of economy and is growing quickly in the last three decades. Compared to manufacturing, operations in the service industry heavily rely on managerial
insights and on-the-fly decision making. There is a strong need to leverage operations research techniques, such as big data analytics, mathematical modeling, and real-time decision support, to reduce cost, improve service quality, and increase revenue. This article provides a multi-objective mathematical model that can improve customer service by minimizing service time, and reduce cost by minimizing the number of employees while maintaining an acceptable level of customer service. The case study of the mathematical model is applied to a quick-service restaurant, but may be extended to many applications in the service industry to improve customer service and reduce cost. The mathematical model uses historical demand data and determine for the next planning horizon the total number of employees required, how they are assigned to different workstations, and expected service times for customers. It takes less than one second to compute an efficient point to the multi-objective model, which can be used for real-time decision support in the service industry.

The mathematical model is robust and provides optimal employee assignment and service rates for workstations in many situations. Managers may determine the importance of different objectives and the model is then applied to provide an optimal solution within one second. Alternatively, some managers prefer to have a variety of options. The model can produce pareto optimal solutions or an efficient frontier with multiple efficient points. Managers can choose the solution that best reflects long-term corporate visions and short-term priorities. For example, senior managers prefer to have fewer employees and cut down the cost since their income is tied to company profit (Hur et al., 2004) whereas the company (or corporate) as a whole aims to have an acceptable service time. The mathematical model can provide much needed information for managers with respect to the tradeoff between multiple conflicting objectives.
More and more companies in the service industry are incorporating automation as part of their workforce (Tanizaki et al., 2017). The service industry needs highly skilled employees for complex tasks (e.g., tasks that involve interactions with customers), and cannot replace all employees with robots. A workforce that combines automation and humans will exist for the foreseeable future. The schedule of tasks and assignment of both robots and humans is a great challenge. Future research may target this promising area and expand the mathematical model developed in this article for applications the integrate humans and artificial intelligence.

References
Azizi, N., Zolfaghari, S., and Liang, M., 2010, “Modeling job rotation in manufacturing systems: The study of employee’s boredom and skill variations,” International Journal of Production Economics, 123, 69-85.
Bureau of Economic Analysis, 2001, “Gross domestic product by industry,” U.S. Department of Commerce.
Choi, K., Hwang, J., and Park, M., 2009, “Scheduling restaurant workers to minimize labor cost and meet service standards,” Cornell Hospitality Quarterly, 50(2), 155-167.
Chua, P. C., Wirawan, H. T., and Chua, T. J., 2017, “Multi-skilled manpower scheduling with part-time consideration: Case study,” in Proceeding of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore.
Corominas, A., Pastor, R., and Rodriguez, E., 2006, “Rotational allocation of tasks to multifunctional workers in a service industry,” International Journal of Production Economics, 103, 3-9.
De Bruecker, P., Beliën, J., Van den Bergh, J., and Demeulemeester, E., 2018, “A three-stage mixed integer programming approach for optimizing the skill mix and training schedules for aircraft maintenance,” *European Journal of Operational Research*, 267(2), 439-452.

Gattis, J. L., Zaman, N., Tauxe, G. W., and Marshment, R. S., 1992, “Analyzing fast-food drive-up window site impacts,” *Site Impact Traffic Assessment: Problems and Solutions*, 16-20.

Hopp, W. J., Iravani, S. M. R., Yuen, G. Y., 2007, “Operations systems with discretionary task completion,” *Management Science*, 53(1), 61-77.

Hur, D., Mabert, V. A., and Bretthauer, K. M., 2004, “Real-time work schedule adjustment decisions: An investigation and evaluation,” *Production and Operations Management*, 13(4), 322-339.

Kwan, S. K., Davis, M. M., and Greenwood, A. G., 1988, “A simulation model for determining variable worker requirements in a service operation with time-dependent customer demand,” *Queueing Systems*, 3(3), 265-275.

Mac-Vicara, M., Ferrera, J. C., Muñoz, J. C., and Henao, C. A., 2017, “Real-time recovering strategies on personnel scheduling in the retail industry,” *Computers & Industrial Engineering*, 113, 589-601.

Marynissen, J. and Demeulemeester, E., 2019, “Literature review on multi-appointment scheduling problems in hospitals,” *European Journal of Operational Research*, 272, 407-419.

Shimmura, T., Akamatsu, M., Takenaka, T., and Oura, S., 2011, “Study of process improvement in restaurant by using cooking behavior analysis and customer’s order information,” *Journal of Japan Industrial Management Association*, 62, 12-20.

Tan, T. F. and Netessine, S., 2014, “When does the devil make work? An empirical study of the impact of workload on worker productivity,” *Management Science*, 60(6), 1574-1593.
Tanizaki, T., Shimmura, T., and Fujii, N., 2017, “Shift scheduling to improve customer satisfaction, employee satisfaction and management satisfaction in service workplace where employees and robots collaborate,” in Proceedings of the 5th International Conference on Serviceology for Services, Vienna, Austria, 15-25.

Vermuyten, H., Namorado Rosa, J., Marques, I., Beliën, J., and Barbosa-Póvoa, A., 2018, “Integrated staff scheduling at a medical emergency service: An optimisation approach,” Expert Systems with Applications, 112, 62-76.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- EAPMApendixA2021.docx