

A Markov-Based Update Policy for Constantly Changing Database Systems

Wei Zong, Feng Wu, and Zhengrui Jiang

Abstract—In order to maximize the value of an organization's data assets, it is important to keep data in its databases up-to-date. In the era of big data, however, constantly changing data sources make it a challenging task to assure data timeliness in enterprise systems. For instance, due to the high frequency of purchase transactions, purchase data stored in an enterprise resource planning system can easily become outdated, affecting the accuracy of inventory data and the quality of inventory replenishment decisions. Despite the importance of data timeliness, updating a database as soon as new data arrives is typically not optimal because of high update cost. Therefore, a critical problem in this context is to determine the optimal update policy for database systems. In this study, we develop a Markov decision process model, solved via dynamic programming, to derive the optimal update policy that minimizes the sum of data staleness cost and update cost. Based on real-world enterprise data, we conduct experiments to evaluate the performance of the proposed update policy in relation to benchmark policies analyzed in the prior literature. The experimental results show that the proposed update policy outperforms fixed interval update policies and can lead to significant cost savings.

Index Terms—Data quality, data timeliness, enterprise resource planning (ERP), Markov decision process, update policy.

I. INTRODUCTION AND MOTIVATION

KEEPING data up-to-date in a database system is critical in order to maximize the value of an organization's data assets. However, constantly changing data sources often make it a challenging task to assure data timeliness in enterprise database systems. Because enterprise resource planning (ERP) systems are one of the most widely adopted types of software systems in today's organizations [1], [2], we use ERP systems to illustrate the challenges, the solutions, and the benefits of achieving data timeliness in a modern database system.

By integrating processes and data from internal functions of an organization and other sources, a successful ERP system can

significantly improve the organization's efficiency, productivity, and competitiveness in the marketplace [3], [4]. However, the advantages that an ERP provides for an organization depend on the quality of data it processes. In fact, data quality has been found to be one of the critical factors in the successful operation of ERP systems [5], [6].

In the era of big data, data often arrive from multiple sources at a high velocity, which brings great challenges to data management [7], [8]. Under this circumstance, organizations hope their data can be updated quickly to capture the data changes, so that at the time of application, current and high quality data can be extracted to support effective decision-making. In fact, data timeliness is identified as one of the most important data quality dimensions for database systems [9], [10]. This is because only when the data in an ERP system is kept current, effective decisions can be made. If the data in a database system is out of date, decisions made based on obsolete data can be incorrect or ineffective, leading to possible financial losses. Take the inventory data in an ERP system as an example, if a new purchase order has been filled, but the new purchase data are not updated in the system in a timely manner, then the company may incur data staleness cost as a result of inefficient inventory replenishment decisions made based on the inaccurate and stale inventory data.

An obvious solution for such data staleness problem is to update the database as soon as a change in the data source occurs, but such a policy is typically inefficient or even infeasible in practice, because it is very costly to keep the data in a database system up-to-date in real time. Sometimes, the database may be under heavy use to support business operations, hence the update has to be delayed. There are also scenarios where database update often cannot be completed without human intervention. In the latter case, the personnel cost could be much higher than the equipment and computation cost [11]. Furthermore, updating information systems frequently can affect system availability [12]. Therefore, to decide when and how to update data in a database system, we need to carefully construct analytical models to balance the tradeoff between staleness cost and update cost.

The remainder of the paper is organized as follows. Section II provides a brief introduction to the data timeliness problem and an overview of the main research efforts on data update policies in the prior literature. Section III introduces the purchase data update problem and the analytical models. The analysis of the optimal update policy and computation method is provided in Section IV. Section V reports on the experiments conducted to evaluate the effectiveness of the proposed update policy in relation to benchmark policies. Finally, Section VI presents

Manuscript received April 29, 2016; revised August 11, 2016; accepted December 24, 2016. This work was supported by the National Natural Science Foundation under Grant 71428003, Grant 71471144, and Grant 71071126, and by the Key Lab of the Ministry of Education for Process Control, and the Efficiency Engineering and Experiment Center for the Teaching of Management Education at Xi'an Jiaotong University. Review of this manuscript was arranged by Department Editor Dr. Bin Jiang.

W. Zong is with the School of Economics and Management, Xidian University, Xi'an 710071, China, and also with the School of Management, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wei.zong@stu.xjtu.edu.cn).

F. Wu is with the School of Management, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: fengwu@mail.xjtu.edu.cn).

Z. Jiang is with the College of Business, Iowa State University, Ames, IA 50011 USA (e-mail: zjiang@iastate.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TEM.2017.2648516

0018-9391 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

some concluding remarks and discusses possible future research directions.

II. RELATED LITERATURE

This section first provides an introduction to how data timeliness is defined in the extant literature, followed by a review of data update policies proposed in prior studies.

A. Current Research on Data Timeliness

Data timeliness has long been considered an important data quality dimension and needs improvement for effective decision-making, especially in the age of big data [13]. Prior research has tried to define data timeliness and proposed various timeless metrics in different applications. Wand and Wang [14] defined timeliness as the delay between a change of a real-world state and the resulting modification of the recorded state in an information system. Similarly, Lee and Strong [15] considered timeliness the extent to which the data were sufficiently up-to-date for the task at hand. Ballou *et al.* [16] provided a more general and pragmatic definition and employed currency and volatility to judge whether the data were out of date. Currency is the difference between the time when data change in the real world and the time when users use it for a certain task. Volatility refers to the average time interval when data remain valid.

Based on the definitions in the aforementioned studies, we can see that data timeliness is affected by three important time instances, i.e., the time when data change in the real world, the time when the changes are extracted and recorded in an information system, and the time when the data are delivered to the information users. The entire process and related events affecting data timeliness are depicted in Fig. 1.

Based on the events affecting data timeliness shown in Fig. 1, the literature in the area of data timeliness can be classified into two research streams. The first research stream mainly focuses on improving data timeliness by quickly extracting and recording the data changes in the real world into an information system, i.e., the data update policy for databases. The second research stream focuses on improving data timeliness by quickly delivering the data changes or the requested information quickly to information users. The second stream features mainly study on how to improve the efficiency of update policies [10], [17], [18], query processing algorithms [9], [19], [20], and system response speed [21].

In this paper, the research problem we focus on designing optimal data update policies to achieve data timeliness. Data timeliness in this paper refers to the difference between the time when data change in the real world and the time when these changes are input into a database system, which corresponds to the first two events in Fig. 1. Therefore, our research belongs to the first research stream and a more detailed review of the related literature on update policy is provided next.

B. Current Research on Data Update Policy

The process of updating the data contents in information systems to reflect the changes occurring at the data sources is

termed as synchronization or materialized view maintenance in some prior studies. Chandy *et al.* [22] presented models and techniques to determine the optimal time interval for database checkpoints. Srivastava and Rotem [23] focused on the efficiency of data update operation and obtained the optimal update frequency based on a stochastic model by minimizing a linear weighted sum of the query waiting cost and processing cost. However, the data staleness cost was not explicitly considered in their research. Segev and Fang [24] designed and proposed the optimal time-based and query-based data update policies, respectively. The objective of their models was to minimize the currency on user queries rather than minimizing the data staleness cost. According to the different properties of updates and database views, Adelberg *et al.* [25] proposed some data update strategies by analyzing various properties of updates and views to balance the transaction deadlines with database currency. Ling and Mi [26] and Dey *et al.* [27] discussed a time-based update policy and designed the optimal update frequency by minimizing the total data staleness cost and synchronization cost. Mannino and Walter [28] explored the short-term and long-term influences of data refresh policies on the traditional information system success measures by conducting a field study. However, a common characteristic of these studies is that the optimal data update policies derived are mainly based on a fixed interval. That is to say, the database is updated after a fixed time interval, e.g., one day/week after a fixed number of user queries have been received by a database, or after a fixed number of new data changes have arrived at a database. To increase the flexibility of data update process, we propose an aperiodic data update policy, illustrated using purchase data update in ERP systems, and test its effectiveness and optimality by comparing it with traditional fixed interval data update policies. Although aperiodic policies have been studied in [11] and [29], the former study focuses on refreshing the knowledge learned from data (e.g., association rules computed from sales data) instead of updating the data itself, and the decision variables, decision models, and cost function considered are also different from ours, leading to different update policies.

III. DESCRIPTION ON RESEARCH PROBLEM AND MODEL

Although the database update policy proposed in this study is applicable in a wide range of data and systems, as explained earlier, for ease of exposition, we use purchase data in an ERP system to motivate the problem and illustrate the model and solutions.

A. Purchase Data Update Process in ERP Systems

Purchase data play an important role in decision-making in manufacturing companies. It is used not only for studying past purchase patterns, but also for updating inventory status of materials and products. For ease of illustration, we consider one type of information inquiry, i.e., inventory inquiry, to an ERP system. User inquiries of inventory status for materials are frequently posted to the ERP system to check the current inventory level and support inventory replenishment decisions. Due to

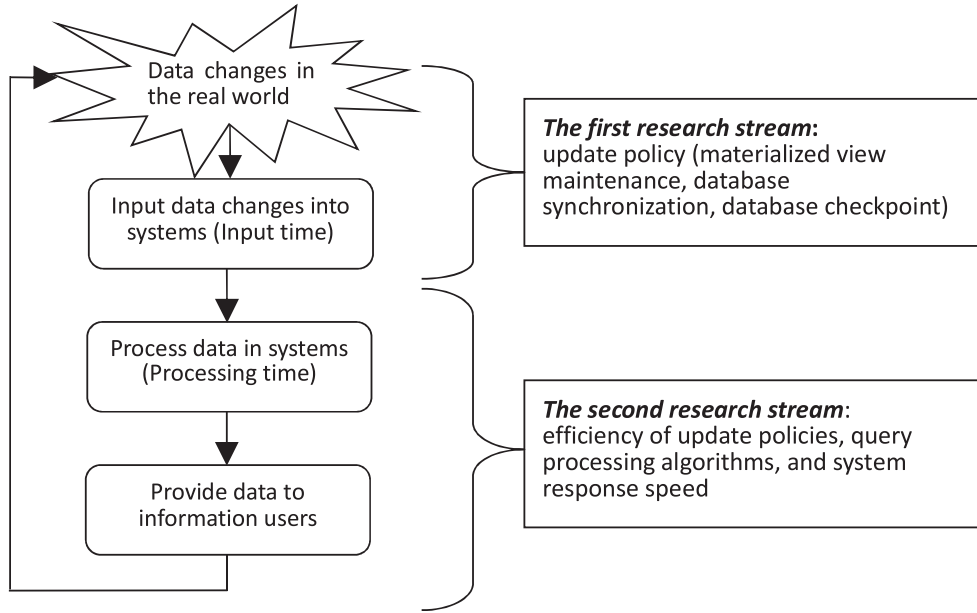


Fig. 1. Events and classification of prior research on data timeliness.

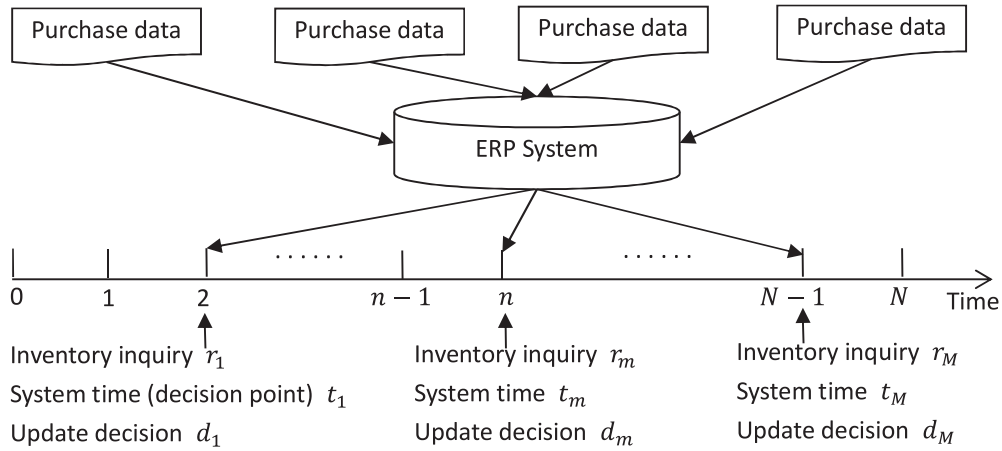


Fig. 2. Timing of purchase data update decisions.

the continuously arriving purchase data, the purchase and inventory data recorded previously may no longer be valid and accurate when an inventory inquiry arrives. Therefore, the purchase data in the ERP system need to be updated in a timely manner to reflect the changes occurring at purchase data sources and provide more accurate data to support inventory replenishment decisions. If the ERP system is not updated with the new purchase data, the inventory inquiry will receive outdated data, which can result in suboptimal decisions and possible financial losses to the company. Conversely, updating the ERP system immediately on the arrival of new purchase data can ensure the timeliness and freshness of the data, but the resulting frequent updates lead to higher update cost. Therefore, a decision regarding whether or not to update the ERP system has to be made when (or just before) an inventory inquiry arrives at the ERP system in a way that minimizes the sum of staleness cost and update cost. The timing of purchase data update decisions can be illustrated in Fig. 2.

B. Definition and Description on the Research Model

In Fig. 2, the data update decisions are discrete, i.e., the data update decision has to be made at each decision point, and each data update decision is only determined by the current system state instead of the previous system states. This conforms to the characteristics of Markov decision process. Therefore, we model the purchase data update problem in ERP as a Markov decision process. System state and transition probability are two indispensable parameters in a Markov decision process. Therefore in this section, we start with the definitions and analysis of system state and transition probability, followed by the derivation of system cost and objective function. The key notations used in this paper are listed in Table AI in the Appendix.

C. Analysis of System State and Transition Probability

Many prior studies have assumed that information inquiries and data updates to a database approximately follow a Poisson

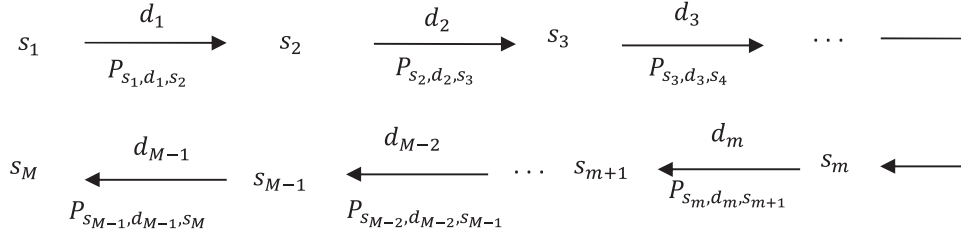


Fig. 3. Process of system state transition.

process [26], [27], [30]. Following the prior literature, we also assume that the arrival of new purchase data to an ERP system is characterized by a Poisson process with an intensity rate of λ_u and that inventory inquiries arriving to an ERP system follow another independent Poisson process with an intensity rate of λ_r . The expected number of inventory inquiries submitted to an ERP system in a predetermined time horizon is denoted by M . The system time t_m refers to the time when the m th inventory inquiry arrives at an ERP system, where $m = 1, 2, \dots, M$. t_m can also be considered the decision points because an update decision d_m needs to be made when the m th inventory inquiry arrives at an ERP system. The update decision d_m can take one of the two possible values, 0 and 1. When $d_m = 0$, it means that the ERP system is not updated. When $d_m = 1$, it indicates that the ERP system is updated with the accumulated purchase data incorporated into the system. As expected, the value of d_m is determined by the system state s_m . In this study, we define system state s_m as the quantity of unprocessed purchase records accumulated by time t_m , where a purchase record describes the purchase details for a particular type of material, such as the material name, purchase date, purchase quantity, purchase price, etc. The decision made at system time t_m will affect the system state at the next system time t_{m+1} . Due to the different values of d_m , the system state at the next decision point should be examined separately.

If $d_m = 0$, it means that the ERP system is not updated at system time t_m . As a result, the accumulated and unprocessed purchase data at time t_m will be carried over to the next system time t_{m+1} . Conversely, if $d_m = 1$, the ERP system is updated at system time t_m . Then, the accumulated purchase data at time t_m will be integrated into the ERP system instead of being carried over to the next system time t_{m+1} . Therefore, the system state s_{m+1} with different update decisions can be represented by

$$s_{m+1} = \begin{cases} s_m + I_{m,m+1} & d_m = 0, \\ I_{m,m+1} & d_m = 1. \end{cases} \quad (1)$$

where $I_{m,m+1}$ is the quantity of purchase data accumulated from time t_m to t_{m+1} .

The system transition from the current system state s_m to the next system state s_{m+1} is captured by the transition probability $P_{s_m,d_m,s_{m+1}}$. As illustrated in Fig. 2, when the first inventory inquiry arrives at the ERP system at system time t_1 , an update decision d_1 needs to be made under the current system state s_1 . After that, the ERP system stochastically transits to the next state s_2 with a transition probability P_{s_1,d_1,s_2} . At the system state s_2 , a decision d_2 is made and the ERP system stochastically transits

to the next state s_3 with a transition probability P_{s_2,d_2,s_3} , and so on. Fig. 3 illustrates the process of system state transition during purchase data update in ERP systems.

In the above analysis, we have assumed that the arrival of inventory inquiries follows a Poisson process with an intensity rate of λ_r . Under this assumption, the time interval between two successive inventory inquiries follows the exponential distribution with an intensity rate of λ_r . Therefore, the probability of the data quantity accumulated from time t_m to t_{m+1} , i.e., $P(I_{m,m+1} = h)$, takes the form

$$P(I_{m,m+1} = h) = \int_0^\infty P(I_{m,m+1} = h | (t_{m+1} - t_m)) \cdot f(t_{m+1} - t_m) d(t_{m+1} - t_m) \quad (2)$$

where $f(t_{m+1} - t_m)$ is the probability density function of time interval $(t_{m+1} - t_m)$ and $f(t_{m+1} - t_m) = \lambda_r e^{-\lambda_r \cdot t}$. Let $(t_{m+1} - t_m) = t$, then (2) can be rewritten as

$$\begin{aligned} P(I_{m,m+1} = h) &= \int_0^\infty P(I_{m,m+1} = h | t) \cdot f(t) d(t) \\ &= \int_0^\infty \frac{e^{-\lambda_u \cdot t} \cdot (\lambda_u t)^h}{h!} \cdot \lambda_r e^{-\lambda_r \cdot t} dt = \frac{(\lambda_u)^h \cdot \lambda_r}{(\lambda_u + \lambda_r)^{h+1}}. \end{aligned} \quad (3)$$

There are two possible values for d_m ($d_m \in \{0, 1\}$), therefore the value of $P_{s_m,d_m,s_{m+1}}$ should be discussed separately.

When $d_m = 0$, from (1) we know that

$$I_{m,m+1} = s_{m+1} - s_m \quad (4)$$

Replacing parameter h in (3) with (4), we get

$$\begin{aligned} P_{s_m,d_m=0,s_{m+1}} &= P(I_{m,m+1} = s_{m+1} - s_m) \\ &= \frac{(\lambda_u)^{(s_{m+1}-s_m)} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(s_{m+1}-s_m+1)}} \end{aligned} \quad (5)$$

When $d_m = 1$, according to (1), we have

$$I_{m,m+1} = s_{m+1}. \quad (6)$$

In this case, substituting h in (3) with (6) yields

$$P_{s_m,d_m=1,s_{m+1}} = P(I_{m,m+1} = s_{m+1}) = \frac{(\lambda_u)^{s_{m+1}} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(s_{m+1}+1)}}. \quad (7)$$

D. Analysis of System Cost

Due to the different cases of update decision made at system time t_m , the system cost $c_m(s_m, d_m)$ should also be discussed

separately. If $d_m = 0$, i.e., the ERP system is not updated at system time t_m , hence the data obtained to answer the inventory inquiry r_m is stale and the staleness cost is incurred, as represented by $c_s(s_m)$ in (8). On the other hand, if $d_m = 1$, i.e., the ERP system is updated at system time t_m , the staleness cost will be avoided but the update cost is incurred, as represented by c_u in (8). In line with prior research (e.g., [27]), we assume that the update cost is a constant and independent of the amount of newly arriving data [27]. In this way, the system cost $c_m(s_m, d_m)$ at t_m with different update decisions can be defined by the following equation:

$$c_m(s_m, d_m) = \begin{cases} c_s(s_m) & d_m = 0 \\ c_u & d_m = 1. \end{cases} \quad (8)$$

Based on the update decision and the corresponding cost at system time t_m , we can formulate the objective function of the purchase data update problem as

$$C = \min E \left(\sum_{m=1}^M (d_m \cdot c_u + (1 - d_m) \cdot c_s(s_m)) \right). \quad (9)$$

For the ease of analysis, (9) can be rewritten as

$$C = \min E (c_1(s_1, d_1) + c_2(s_2, d_2) + \dots + c_M(s_M, d_M)). \quad (10)$$

The objective of problem (10) is to minimize the expected total system cost in a time horizon by finding the optimal sequence of update decisions δ^* at all decision points, i.e., $\delta^* = (d_1^*, d_2^*, \dots, d_M^*)$. If we consider one decision point at a time, the optimal update decision d_m^* at time t_m can be decided by comparing $c_s(s_m)$ with c_u . If $c_s(s_m) \geq c_u$, it is better to update the ERP system and therefore $d_m^* = 1$. If $c_s(s_m) < c_u$, the smaller data staleness cost indicates that there is no need to update the ERP system, therefore $d_m^* = 0$. However, with all decision points considered, the above-mentioned simple method is generally not optimal. Therefore, we need to develop an efficient method to obtain the optimal update sequence for the entire time horizon.

E. Analysis of Staleness Cost Function

Staleness cost $c_s(s_m)$ reflects the effect of accumulated unprocessed purchase data on future purchase decisions. In an ERP system, if an inventory inquiry is answered by the stale data, the purchase decisions based on those outdated inventory data will likely be inefficient and inaccurate, leading to the unnecessary purchase cost c_p and inventory cost c_I . Therefore, in the present research the staleness cost $c_s(s_m)$ is defined as

$$c_s(s_m) = (c_p + c_I) \cdot F(s_m) \quad (11)$$

where purchase cost c_p and inventory cost c_I are considered constant. $F(s_m)$, which we refer to as the *staleness severity function*, represents the severity of inefficiency or financial loss as a result of data staleness in the ERP systems and is a function of system state s_m . Note that similar concept of staleness severity has been proposed in prior research (e.g., [26], [27]). Recall that s_m represents the accumulated quantity of unprocessed purchase data at time t_m , therefore, we can see from (11)

that the staleness cost is dependent on the quantity of accumulated unprocessed purchase data at t_m . The larger the amount of unprocessed purchase data at t_m , the more likely the focal company will make inefficient inventory replenishment decisions, thus resulting in a higher staleness cost. Therefore, $c_s(s_m)$ is a monotone increasing function in system state s_m . Furthermore, we define $F(s_m) = 0$ when $s_m = 0$. Then, we have $c_s(s_m) = 0$, indicating that if the quantity of unprocessed purchase data accumulated by time t_m is zero, i.e., the purchase data in ERP system is up-to-date at t_m , then there will be no stale data and the staleness cost is zero.

By analyzing the characteristics of (10), we can find that the staleness severity function $F(s_m)$ plays a critical role in determining the value of the objective function. Theorem 1 describes the relationship between the optimal expected total system cost and the staleness severity function.

Theorem 1: For a same system state s_m , let C_1 and C_2 denote the optimal expected total system cost corresponding to staleness severity $F_1(s_m)$ and $F_2(s_m)$. If $F_1(s_m) \geq F_2(s_m)$, then $C_1 \geq C_2$, where $s_m \in S$ and $m = 1, 2, \dots, M$. (All proofs are provided in the Appendix.)

As explained earlier, $F(s_m)$ represents the severity of inefficiency or financial loss as a result of data staleness. Based on Theorem 1, for the same system state s_m , a higher value of $F(s_m)$ indicates that the cost of stale data is higher. This shows from another perspective that the company with a higher staleness severity function is more sensitive to stale data, and therefore may put more emphasis on maintaining data timeliness for decision-making.

IV. ALGORITHM FOR DERIVING THE OPTIMAL UPDATE POLICY

In order to obtain the minimal expected total cost as specified in the objective function of problem (10), we need to examine all system state s_m at each decision point d_m for $m = 1, 2, \dots, M$. This seems impossible because the number of possible values of system state s_m is infinite due to the stochastic nature of the accumulated purchase data at t_m . Fortunately, based on the above analysis we can see that each data update decision at each decision point is only determined by the current system state instead of the previous system states. Therefore, the entire update decision problem can be formulated as a Markov decision process. In this section, we analyze the properties of the objective function and the optimal update policy based on the Markov decision process. Then, we propose a computation algorithm to derive the optimal update policy δ^* using a backward induction method.

A. Analysis of the Optimal Update Policy

Based on (10) and the theory of Markov decision process, we let

$$V_1 = E(c_1(s_1, d_1) + c_2(s_2, d_2) + \dots + c_M(s_M, d_M))$$

represent the expected total system cost (value function) from time t_1 to time t_M . Similarly

$$V_2 = E(c_2(s_2, d_2) + \dots + c_M(s_M, d_M))$$

represent the expected total system cost from time t_2 to time t_M . Thus, we have

$$V_1 = c_1(s_1, d_1) + V_2$$

and more generally

$$V_m = c_m(s_m, d_m) + V_{m+1}.$$

That is to say, the value function V_m at system time t_m can be considered the sum of the system cost at current decision point t_m and the value function at the next decision point t_{m+1} . When $m = M$, let $V_{M+1} = 0$ and we get

$$V_M = c_M(s_M, d_M).$$

Because the accumulated purchase data arriving to an ERP system is stochastic, all the possible values of accumulated purchase data quantity at each decision point should be considered. Therefore, we include the probability distribution of system state s_m transiting to s_{m+1} in the value function. Then, $V_m = c_m(s_m, d_m) + V_{m+1}$ can be rewritten as

$$V_m = c_m(s_m, d_m) + \sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot V_{m+1} \quad (12)$$

where V_m represents the expected total system cost from system time t_m to t_M .

Based on the principle of Markov decision process, to obtain the minimal expected total system cost V_m^* , we must find a set of optimal update decision δ_m^* under the current system state s_m that can minimize V_m , where $\delta_m^* = (d_m^*, d_{m+1}^*, \dots, d_M^*)$ and $m = 1, 2, \dots, M$. The optimal expected total system cost V_m^* from system time t_m to t_M is defined as

$$V_m^* = \min \left(c_m(s_m, d_m) + \sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot V_{m+1}^* \right). \quad (13)$$

When $m = M$, we can get $V_M^* = c_M(s_M, d_M)$ because $V_{M+1} = 0$.

From the above analysis, we know that V_1^* is the minimal expected total system cost from system time t_1 to t_M . Considering that there are infinite number of possible system states at the start decision point t_1 , the optimal expected total system cost for all M inventory inquiries in a time horizon can be represented by

$$C = \sum_{s_1 \in S} V_1^* \cdot P_{s_1} \quad (14)$$

where P_{s_1} is the probability that the system state is s_1 at time t_1 .

As mentioned before, the system state s_1 is represented by the quantity of accumulated unprocessed purchase data by time t_1 . Recall that t_1 is the time that the first inventory inquiry arrives at the ERP system. Therefore, the data quantity accumulated by time t_1 equals the data quantity accumulated from the start time of the time horizon to time t_1 . According to (3), P_{s_1} can be represented by

$$P_{s_1} = P(I_{0,1} = q) = \frac{(\lambda_u)^q \cdot \lambda_r}{(\lambda_u + \lambda_r)^{q+1}}. \quad (15)$$

In order to obtain the optimal update policy from system time t_m to t_M , we next analyze the characteristics of V_m^* and V_m . Based on the optimal expected total system cost V_m^* shown in (13), we define

$$R(s_m, d_m) = c_m(s_m, d_m) + \sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot V_{m+1}^*. \quad (16)$$

Then, (13) can also be rewritten as

$$V_m^* = \min R(s_m, d_m).$$

One of the very important properties of Markov decision process is the optimality of monotone policies. It is obtained for the problem of maximizing an objective function. However, the objective function in this study is to minimize the expected total system cost. Therefore, we redefine our problem and transform the minimum objective function (13) into a maximum one

$$G_m^* = \max \left(-c_m(s_m, d_m) + \sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot G_{m+1}^* \right) \quad (17)$$

where G_m^* is the maximal reward from time t_m to t_M and $G_m^* = -V_m^*$.

Based on the optimality of monotone policies in Markov decision process [31], we obtain the following properties for problem (17).

Theorem 2: For problem (17), let $m = 1, 2, \dots, M-1$, if $-c_m(s_m, d_m)$ is nonincreasing in s_m for all $d_m \in D$, $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}}$ is nondecreasing in s_m for all $s_{m+1} \in S$ and $d_m \in D$, $-c_m(s_m, d_m)$ is a superadditive function on $S \times D$, $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot G_{m+1}^*$ is a superadditive function on $S \times D$ for all $s_{m+1} \in S$, and $-c_M(s_M, d_M)$ is nonincreasing in s_m , then the optimal decision policy d_m^* is monotone nondecreasing in system state s_m for $m = 1, 2, \dots, M$.

We next provide the definition on superadditive (see Definition 1) and prove that each of the condition is required to get the monotone nondecreasing optimal policy. Please refer to the Appendix for the detailed proof of Theorem 2.

DEFINITION 1: A real-valued function $g(x, y)$ on $X \times Y$ is superadditive if for $x^+ \geq x^-$ in X and $y^+ \geq y^-$ in Y

$$g(x^+, y^+) + g(x^-, y^-) \geq g(x^+, y^-) + g(x^-, y^+). \quad (18)$$

From (13) and (16), we can see that the optimal expected total system cost V_m^* can be interpreted as the comparison result between the expected total system cost under an update decision $R(s_m, d_m = 1)$ and the expected total system cost under a no update decision $R(s_m, d_m = 0)$. If $R(s_m, d_m = 0)$ at system time t_m is equal to or greater than $R(s_m, d_m = 1)$, then updating the ERP system with the newly arriving purchase data is a better choice. Otherwise, it is better not to update the ERP system.

From Theorem 2, we conclude that the optimal update policy d_m^* is monotone nondecreasing in system state s_m . We already know that d_m is composed by 0 and 1. Therefore, monotone nondecreasing of d_m^* means that d_m^* is increasing from 0 to 1. That is to say, there is a control limit l_m^* for system state s_m at each decision point determining the optimal update policy is

0 or 1. The optimal update policy can be transformed into the form

$$d_m^* = \begin{cases} 0 & s_m < l_m^* \\ 1 & s_m \geq l_m^* \end{cases} \quad (19)$$

Based on (19), if the accumulated purchase data s_m are less than the control limit l_m^* , it is optimal not to update the system, and when the accumulated purchase data are equal to or greater than l_m^* , it is optimal to update the system. Therefore, l_m^* is the minimal quantity of accumulated purchase data satisfying $R(s_m, d_m = 0) \geq R(s_m, d_m = 1)$ and $l_m^* \in Z^+$. In this way, the problem of finding the optimal update policy δ^* is reduced to that of determining the control limits l^* , where $l^* = (l_1^*, l_2^*, \dots, l_m^*, \dots, l_M^*)$.

B. Computational Method

Backward induction is an efficient method for solving Markov decision problem with a finite time horizon. Therefore, we design an algorithm to find the optimal update policy l^* for the purchase data update problem based on the backward induction method. As explained before, the objective is to determine the control limit l_m^* at each decision point, which is the minimal quantity of accumulated purchase data satisfying $R(s_m, d_m = 0) \geq R(s_m, d_m = 1)$. After further analyzing $R(s_m, d_m)$ in (16), we find that it is impossible to directly calculate $R(s_m, d_m = 0) \geq R(s_m, d_m = 1)$ for $m = 1, 2, \dots, M$, because at each step we need to compute V_{m+1}^* for an infinite number of possible system state s_{m+1} . However, from Theorem 2, we know that the optimal decision policy d_m^* is monotone nondecreasing in system state s_m for $m = 1, 2, \dots, M$. Therefore, we can compare and record the values of $R(s_m, d_m = 0)$ and $R(s_m, d_m = 1)$ at each step when s_m gradually increases from zero until the minimal s_m makes $R(s_m, d_m = 0)$ equal or larger than $R(s_m, d_m = 1)$. In fact, the minimal s_m by solving $R(s_m, d_m = 0) \geq R(s_m, d_m = 1)$ is the optimal control limit l_m^* at each decision point. Furthermore, when $m = M$, we have $R(s_M, d_M) = c_M(s_M, d_M)$, and when $m = M-1, \dots, 2, 1$, we have $R(s_m, d_m) = c_m(s_m, d_m) + \sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot V_{m+1}^*$, therefore we should consider the two different update decisions for $R(s_m, d_m)$ separately.

In backward induction, we first obtain l_M^* , which is the minimal quantity of accumulated purchase data at time t_M that satisfies

$$R(s_M, d_M = 0) \geq R(s_M, d_M = 1) \quad (20)$$

Based on (8) and (16), and $V_{M+1}^* = 0$, (20) is reduced to

$$c_s(s_M) \geq c_u \quad (21)$$

Similarly, for $m = M-1, M-2, \dots, 2, 1$, we need to check whether $R(s_m, d_m = 0) \geq R(s_m, d_m = 1)$ is satisfied as s_m gradually increases from zero, a procedure that helps us obtain l_m^* at each decision point. The complete algorithm is summarized in Fig. A1 in the Appendix.

TABLE I
COMPARISONS OF EXPECTED TOTAL COST UNDER DIFFERENT DATA
UPDATE POLICES

	C	$C_{time}(t')$	$C_{inquiry}(i')$	$C_{update}(u')$
$\lambda = 0.0001$	40 246	53 118 (24.2%)	61 960 (35%)	53 111 (24.2%)
$\lambda = 0.0005$	61 073	118 770 (48.6%)	162 980 (62.5%)	118 740 (48.6%)
$\lambda = 0.001$	67 966	167 970 (59.5%)	256 600 (73.5%)	167 900 (59.5%)
$\lambda = 0.005$	76 544	375 590 (79.6%)	963 900 (92.1%)	375 240 (79.6%)
$\lambda = 0.01$	77 975	530 560 (85.3%)	1 844 000 (95.8%)	529 870 (85.3%)
$\lambda = 0.05$	79 207	1 197 200 (93.4%)	9 021 700 (99.1%)	1 191 000 (93.3%)
$\lambda = 0.1$	79 366	1 807 300 (95.6%)	17 564 000 (99.5%)	1 663 400 (95.2%)
$\lambda = 0.5$	79 491	5 807 400 (98.6%)	73 565 000 (99.8%)	3 424 700 (97.7%)

V. EXPERIMENTS AND RESULTS FOR PERFORMANCE EVALUATION

To test the performance and effectiveness of the aperiodic purchase data update policy obtained from the proposed method, we conducted experiments based on a real-life dataset from the Hongmen Advanced Technology Co., Ltd. (or Hongmen for short) located in the Shen Zhen city in China (<http://www.hongmen.com/>). Hongmen has a history of almost 20 years, a registered capital of 66.8 million RMB, more than 550 employees. It is one of the largest gate manufacturers in China, and specializes in developing, designing, and producing over 21 series of products such as automatic gates, IC card intelligent systems, intelligent parking management systems, etc. By the end of the year 2015, its total sales, cost of goods sold, and net profit have reached RMB257.353 million, RMB165.200 million, and RMB10.846 million, respectively. In May 2011, Hongmen invested more than 2 104 626 RMB to implement the SAP-R/3 ERP system, which was made up 90.98% of its total IT cost. The variety of products in this company require a wide variety of materials be purchased and managed, which brings a big challenge to maintain the timeliness of inventory data in its ERP system. We use parameter values estimated from real-life data from this company in our experiments.

A. Parameter Settings

In this paper, the unit of time is set to be one day and the entire time horizon is one year or 365 days in our experiments. According to the real operation situation in Hongmen, the arrival rate of new purchase data λ_u , representing the average number of purchase records arriving to the ERP system in one day, was set to 182. The inventory state was checked every seven days in the company, therefore the arrival rate of inventory inquiries λ_r was 1/7. Based on information provided by the ERP system manager of Hongmen, the update cost c_u consists of computation cost and personnel cost. Computation cost mainly includes the hardware and software cost incurred when updating the ERP system. According to the ERP system operation manager, the average maintenance cost on ERP hardware and software is RMB474 500 per year, and the average monthly salary for the material manager is RMB6900. Hongmen estimated that the cost of running one update approximately equals one day's of hardware and software maintenance cost and one day's salary of the material

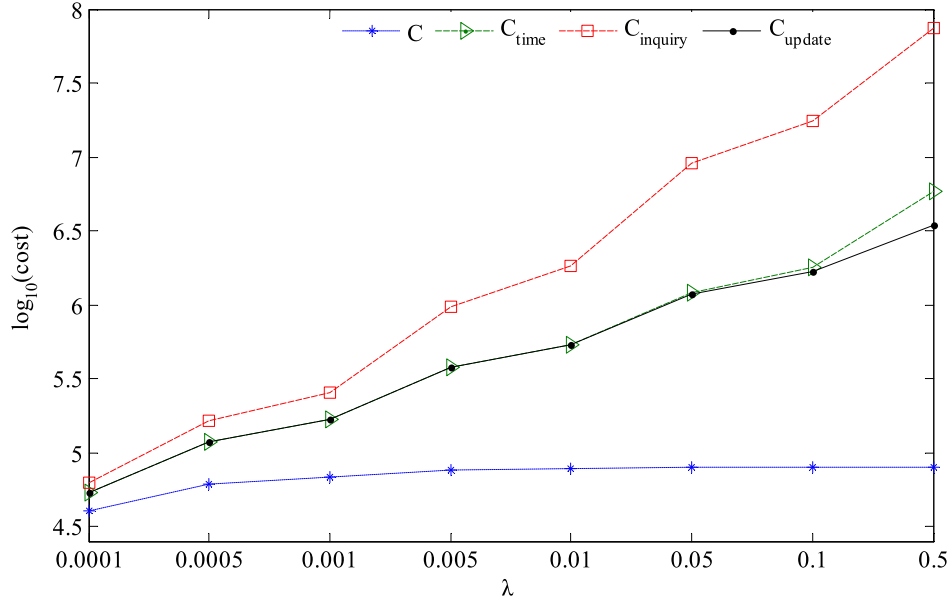


Fig. 4. Logarithm of expected total cost in Table I.

TABLE II
COMPARISONS OF EXPECTED TOTAL COST UNDER DIFFERENT UPDATE
POLICIES AND STALENESS SEVERITY FUNCTIONS

Parameters		C	$C_{time}(t')$	$C_{inquiry}(i')$	$C_{update}(u')$
$F_{logistic}$ $\alpha = 15$	$\beta = 0.001$	7429.6	33 154 (77.6%)	36 598 (79.7%)	33 151 (77.6%)
	$\beta = 0.005$	26 907	74 205 (63.7%)	91 460 (70.6%)	74 191 (63.7%)
	$\beta = 0.01$	39 758	104 940 (62.1%)	139 450 (71.5%)	104 910 (62.1%)
	$\beta = 0.05$	65 862	234 650 (71.9%)	424 850 (84.5%)	234 510 (71.9%)
	$\beta = 0.1$	72 004	335 920 (78.6%)	786 980 (90.9%)	335 640 (78.5%)
	$\beta = 0.5$	77 889	786 320 (90.1%)	3 954 900 (98%)	784 800 (90.1%)
	$\beta = 1$	78 700	1 112 000 (92.9%)	7 829 900 (98.9%)	1 109 000 (92.9%)
$F_{uniform}$	$\delta = 100$	79 714	609 080 (86.9%)	2 404 800 (96.7%)	608 170 (86.9%)
	$\delta = 500$	75 084	272 390 (72.4%)	544 790 (86.2%)	272 210 (72.4%)
	$\delta = 1000$	70 502	192 610 (63.4%)	312 280 (77.4%)	192 520 (63.4%)
	$\delta = 3000$	58 833	111 200 (47.1%)	149 950 (60.8%)	111 170 (47.1%)
	$\delta = 5000$	51 963	86 137 (39.7%)	109 390 (52.5%)	86 119 (39.7%)
	$\delta = 8000$	45 328	68 097 (33.4%)	82 629 (45.1%)	68 086 (33.4%)
	$\delta = 10000$	42 183	60 908 (30.7%)	72 533 (41.8%)	60 899 (30.7%)
	$\delta = 15000$	36 625	49 728 (26.3%)	57 477 (36.3%)	49 722 (26.3%)

manager. Hence, the unit computation cost and personal cost incurred when updating the ERP system were set to RMB1300 (RMB474 500/365 days) and RMB230 (RMB6900/30 days), respectively. Therefore, the total update cost c_u was RMB1530 (RMB1300+ RMB230).

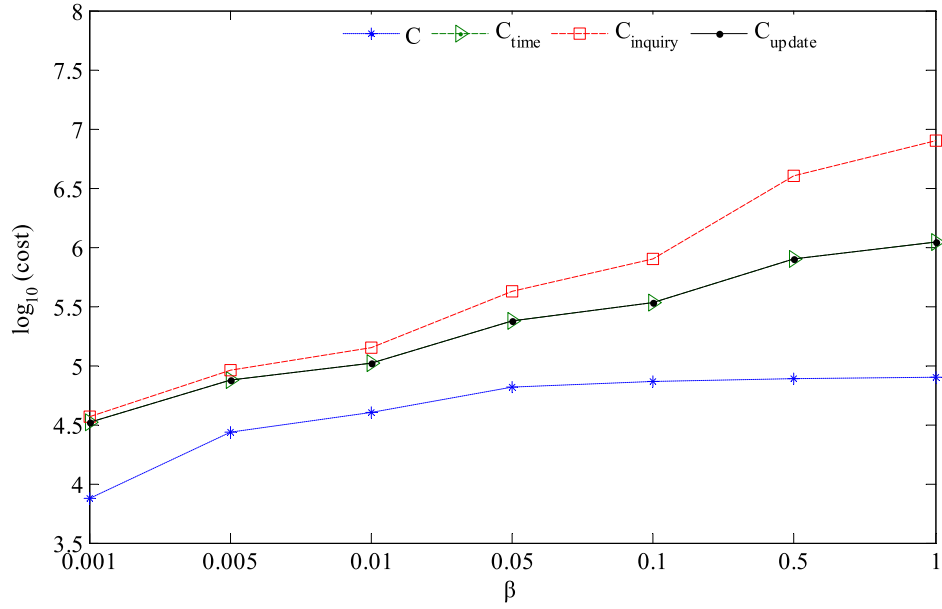
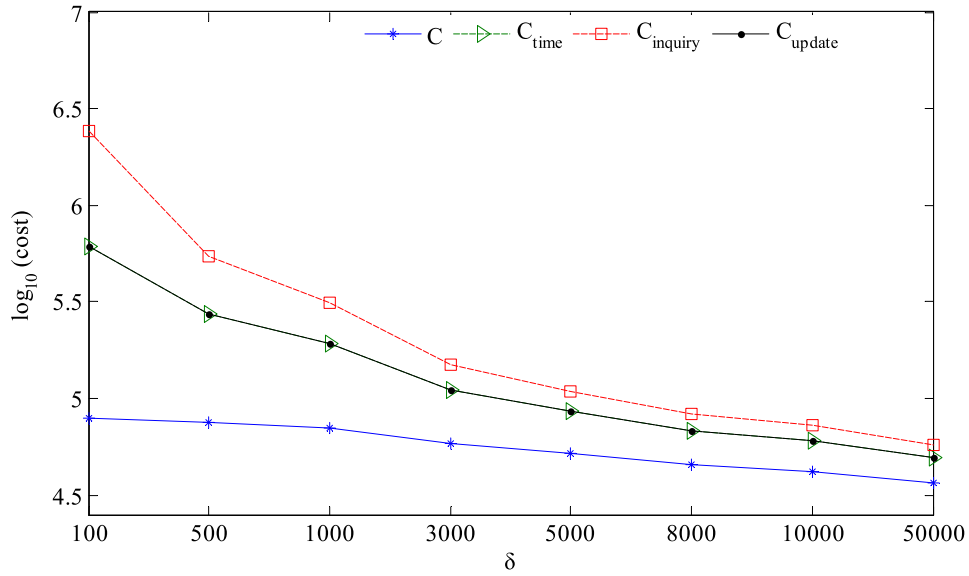
Regarding the staleness cost, according to (11) and the earlier discussion, we know that the staleness cost, which takes the form of $c_s(s_m) = (c_p + c_I) \cdot F(s_m)$, is a monotone increasing function, where c_p and c_I are constants. Therefore, the staleness severity function $F(s_m)$ is also a monotone increasing function of its parameter, the system state s_m . Depending on the real-world problem context, $F(s_m)$ can take different functional forms. For ease of analysis, one possibility is to use the cumulative distribution function (CDF) of a certain distribution to represent the monotonically increasing nature of this function.

In the Hongmen Company, the average purchase cost (setup cost) is RMB1000, mainly including the travel expense, mailing expense, and so on. The ordering policy in the company is similar to the (R, Q) policy, i.e., when the inventory position declines to or below the reorder point R , a batch quantity of size Q is ordered. Based on the calculation from the ERP database in Hongmen, the average purchase quantity Q is 150 units, the average reorder point R is 50 units, and the average holding cost per unit is 20 RMB. Therefore, the average inventory cost is 2500 RMB $((150/2+50)*20)$. In summary, the staleness cost can be rewritten as $c_s(s_m) = 3500 \cdot F(s_m)$.

B. Performance Analysis of the Update Policy

The performance of the optimal purchase data update policy proposed in this paper is evaluated using the cost savings compared with the traditional data update policies based on the fixed interval. The three fixed interval update policies are described as follows [27].

- 1) *Fixed time interval policy*: This can also be called periodical policy, which requires that the information system is updated after a fixed time interval, e.g., every 5 h.
- 2) *Fixed inquiry interval policy*: Under this policy, an information system should be immediately updated when it

Fig. 5. Logarithm of expected total cost with $F_{logistic}$.Fig. 6. Logarithm of expected total cost with $F_{uniform}$.

receives a fixed number of inquiries, e.g., after every five inquiries received by this system.

- 3) *Fixed update interval policy*: This policy demands that an information system is updated after a fixed number of data updates have arrived at the system, e.g., when the number of accumulated new data is equal to or larger than five records.

We define C , C_{time} , $C_{inquiry}$, and C_{update} as the optimal expected total system costs obtained by implementing the aperiodic data update policy proposed in this study, the fixed time interval policy, the fixed inquiry interval policy, and the fixed update interval policy, respectively. Similarly, we let t' , i' , and u' denote the percentages of cost savings achieved by the proposed

aperiodic update policy over the three traditional fixed interval update policies.

In the first set of experimental runs, we let $F(s_m)$ take the functional form of the CDF of exponential distribution (represented by $F_{exponential}$), i.e.

$$F(s_m) = F_{exponential}(s_m) = 1 - e^{-\lambda \cdot s_m} \quad (22)$$

where λ is a parameter of the function. It can be shown that this is a monotone increasing concave function of s_m . To see the performance and advantages of the compared update policies, we conducted a series of experiments while changing the value of the parameter λ . The cost results and the percentage of cost savings achieved by proposed data update policy (in parenthesis)

are summarized in Table I. For ease of comparison, we also plot the logarithms of expected total costs in Fig. 4.

The results in Table I and Fig. 4 show that the proposed aperiodic update policy consistently outperforms all three fixed interval update policies and can lead to significant cost savings. Another apparent result shown in Table I and Fig. 4 is that the optimal expected total system cost increases with the value of parameter λ . This is because the higher value of λ leads to a higher value of $F_{\text{exponential}}(s_m)$ when system state s_m is fixed. According to Theorem 1, the optimal expected total system cost also increases as a result. For the three fixed interval update policies, when the value of parameter λ gets higher, the system becomes more sensitive to the stale data and it has to be updated at a smaller interval. Therefore, the optimal expected total system cost is also increasing with λ . Another noticeable result demonstrated in Table I and Fig. 4 is that among those three fixed interval update policies, the fixed update interval policy is better than the fixed time interval policy, which is in turn better than the fixed inquiry interval policy, i.e., $C_{\text{update}} < C_{\text{time}} < C_{\text{inquiry}}$. This is consistent with the findings of prior research [27].

Additional experiments were conducted to examine the effect of different forms of staleness severity functions on the performances of different update policies. Specifically, in addition to the CDF of exponential distribution, we tried two other monotone increasing functions with different curves. The logistic function (denoted by F_{logistic}) shown in (23) represents a monotone increasing S curve

$$F(s_m) = F_{\text{logistic}}(s_m) = \frac{1}{1 + e^{(\alpha - \beta \cdot s_m)}} \quad (23)$$

The CDF of uniform distribution (denoted by F_{uniform}) is a monotone increasing piecewise linear curve

$$F(s_m) = F_{\text{uniform}}(s_m) = \begin{cases} 0, & s_m < 0 \\ \frac{s_m}{\delta}, & 0 \leq s_m < \delta \\ 1, & s_m \geq \delta. \end{cases} \quad (24)$$

As explained before, we need $F(s_m) = 0$ when $s_m = 0$. Although theoretically the value of α in (23) needs to be infinite to meet this requirement, we set the value of α to 15 because such a value can already make $F(s_m)$ incredibly close to zero (0.0000003059). The comparison of the optimal expected total system costs under four data update policies and changing parameter values in two different staleness severity functions are shown in Table II.

Similar to Fig. 4, we plot the logarithms of expected total costs in Fig. 5 and Fig. 6, corresponding to logistic and uniform functions, respectively.

Once again, the experimental results summarized in Table II and Figs. 5 and 6 show that the update policy proposed in this research consistently and significantly outperforms other three fixed interval update policies. When $F(s_m)$ takes the logistic functional form, the optimal expected total costs of all the four data update policies increase with the value of β . This is because the value of F_{logistic} is higher with a larger β value. When $F(s_m)$ takes the form of the CDF of uniform distribution, the expected total costs of all four data update policies decrease as the δ

value increases. This is because with other factors fixed, the value of $F_{\text{uniform}}(s_m)$ decreases as the value of δ increases. From Table II and Figs. 5 and 6, we also observe that among the three fixed interval update policies, the fixed inquiry interval policy always incurs a higher cost than the fixed time interval policy and the fixed update interval policy, and the cost incurred by the fixed time interval policy is slightly higher than the fixed update interval policy. This result is also consistent with the previous study [27].

VI. CONCLUSION AND DISCUSSIONS

Timeliness is one of the most important quality dimensions of data assets managed by organizations. However, in the era of big data, with the sheer volume of constantly changing data, how to achieve and maintain data timeliness in dynamic database systems has become a challenging yet unresolved operational issue to many organizations. This research tries to address this issue by helping organizations decide when and how frequently to update data in their database systems. We model the data update problem as a Markov decision process and design a dynamic planning algorithm to obtain the optimal update policy based on the backward induction method. A set of experiments, using three different types of staleness severity functions, are conducted to verify the performance of the proposed aperiodic data update policy. The results show that the proposed data update policy consistently and significantly outperforms the traditional fixed interval data update policies.

This study has important practical implications for organizations. First, although fixed interval update policies might be easier to implement, given the amount of cost savings that can be achieved by the proposed aperiodic database update policy, organizations should explore the possibility of using aperiodic update policies to save costs. Second, the staleness severity function and other parameter values have a substantial impact on the optimal update policy. For this reason, efforts should be taken to make sure that the staleness costs are accurately estimated. Third, although we illustrate the proposed update policy using ERP systems, we believe that the aperiodic update policy, with some adaptation, could be used for the maintenance of a wide range of data assets and applications such as patient medical record maintenance, Geographic Information Systems (GIS) data update, website maintenance, etc.

There are a few possible directions for future research, all involving improving or extending the aperiodic policy proposed in the present research. First, the proposed policy is developed for updating reliable and structured data such as purchase data. In other more complex general setting, data could arrive from multiple sources with varying degrees of reliability. One could extend the data update policy proposed in this study for processing and updating data from such more complex data sources. Second, the present study assumes that there is only one type of inquiries, the derived policy could be extended to cover scenarios where there is more than one type of inquiries. Third, it should be possible to derive a data update policy for more general scenarios where the arrival rate of purchase data or

inventory inquiries changes with time, i.e., the parameters λ_u and λ_r are functions of time.

APPENDIX

A. Key Notations

TABLE AI
SUMMARY OF NOTATIONS

Parameter	Description on the parameter
N	The time horizon of the purchase data update problem.
M	The average total number of inventory inquiries arriving at an ERP system in a time horizon.
r_m	The m th inventory inquiry arriving at an ERP system, $m = 1, 2, \dots, M, m \in \mathbb{Z}^+$.
t_m	The time when the m th inventory inquiry arrives at an ERP system.
s_m	The system state at time t_m , representing the accumulated quantity of unprocessed purchase data at time t_m and $s_m \in \mathbb{Z}^+$.
S	The set of system state $s_m \in S$.
D	The decision set.
d_m	The decision made when the m th inventory inquiry arrives at an ERP system at time t_m , $d_m \in D = \{0, 1\}$.
$P_{s_m, d_m, s_{m+1}}$	State transition probability, representing the probability that the system state s_m at time t_m under decision d_m transits to the system state s_{m+1} at time t_{m+1} .
λ_u	The arrival rate of new purchase data in an ERP system.
λ_r	The arrival rate of an inventory inquiry in an ERP system.
$I_{m-1, m}$	The purchase data accumulated between time interval t_{m-1} and t_m .
c_u	The update cost at time t_m .
$c_s(s_m)$	The staleness cost for inventory inquiry r at time t_m as a function of s_m .
$F(s_m)$	The data staleness severity function.
$c_m(s_m, d_m)$	The system cost at time t_m under decision d_m .
C	The optimal expected total system cost for M inventory inquiries in a time horizon.
V_m	The expected total system cost from time t_m to t_M , $V_m \geq 0$.
δ	The update policy in a time horizon, i.e., the update decision made at each decision point, $\delta = (d_1, d_2, \dots, d_M)$.
δ_m^*	The optimal update policy from time t_m to t_M , $\delta_m^* = (d_m^*, d_{m+1}^*, \dots, d_M^*)$.
l_m^*	The optimal control limit at time t_m , $l_m^* \in \mathbb{Z}^+$.
l^*	The optimal control limit for system state, where $l^* = (l_1^*, l_2^*, \dots, l_M^*)$.

B. Proof of Theorem 1

Proof: Let $F_1(s_m) \geq F_2(s_m)$. According to (9) and (10), the optimal expected total system cost can be written as $C = \min E(c_1(s_1, d_1) + c_2(s_2, d_2) + \dots + c_M(s_M, d_M))$.

Let C' and C'' represent the expected total system cost by applying data staleness function $F_1(s_m)$ and $F_2(s_m)$, respectively. $d_m^{*'} = 0$ and $d_m^{*''} = 1$ denote the optimal update policy at time t_m under different staleness functions. Therefore, we can get

$C' = \min E(c_1(s_1, d_1)' + c_2(s_2, d_2)' + \dots + c_M(s_M, d_M)')$ and $C'' = \min E(c_1(s_1, d_1)'' + c_2(s_2, d_2)'' + \dots + c_M(s_M, d_M)'')$.

For $c_1(s_1, d_1)'$ and $c_1(s_1, d_1)''$, when $d_1^{*'} = d_1^{*''} = 0$, according to (8) and (11), we have $c_1(s_1, d_1)' = (c_p + c_I) \cdot F_1(s_m)$ and $c_1(s_1, d_1)'' = (c_p + c_I) \cdot F_2(s_m)$. Therefore, we can get $c_1(s_1, d_1)' \geq c_1(s_1, d_1)''$ for $F_1(s_m) \geq F_2(s_m)$.

When $d_1^{*'} = d_1^{*''} = 1$, according to (8) and (11), we have $c_1(s_1, d_1)' = c_u$ and $c_1(s_1, d_1)'' = c_u$. Therefore, we can get $c_1(s_1, d_1)' = c_1(s_1, d_1)''$ for $F_1(s_m) \geq F_2(s_m)$.

When $d_1^{*'} = 1$ and $d_1^{*''} = 0$, according to (8) and (11), we have $c_1(s_1, d_1)' = c_u$ and $c_1(s_1, d_1)'' = (c_p + c_I) \cdot F_2(s_m)$. $d_1^{*''} = 0$

indicates that the data staleness cost is smaller than the update cost at time t_1 , i.e., $c_1(s_1, d_1)'' = (c_p + c_I) \cdot F_2(s_m) < c_u$. Therefore, we have $c_1(s_1, d_1)' > c_1(s_1, d_1)''$ for $F_1(s_m) \geq F_2(s_m)$.

At last, when $d_1^{*'} = 0$ and $d_1^{*''} = 1$, according to (8) and (11), we can conclude that $c_1(s_1, d_1)' = (c_p + c_I) \cdot F_1(s_m)$ and $c_1(s_1, d_1)'' = c_u$. $d_1^{*'} = 0$ indicates that the data staleness cost is smaller than the data update cost, i.e.

$$(c_p + c_I) \cdot F_1(s_m) < c_u \quad (A1)$$

whereas $d_1^{*''} = 1$ indicates that the data staleness cost is equal to or larger than the data update cost, i.e.

$$(c_p + c_I) \cdot F_2(s_m) \geq c_u. \quad (A2)$$

Based on (A1) and (A2), we know that $F_2(s_m) > F_1(s_m)$, which contradicts the assumption $F_1(s_m) \geq F_2(s_m)$. That is to say, the condition when $d_1^{*'} = 0$ and $d_1^{*''} = 1$ is false and therefore no need to be considered for the next analysis.

Therefore, we can get $c_1(s_1, d_1)' \geq c_1(s_1, d_1)''$ for $F_1(s_m) \geq F_2(s_m)$.

By the above analogy, we can get that $c_2(s_2, d_2)' \geq c_2(s_2, d_2)''$, $c_3(s_3, d_3)' \geq c_3(s_3, d_3)''$ until $c_M(s_M, d_M)' \geq c_M(s_M, d_M)''$. Therefore, $C' \geq C''$ for $F_1(s_m) \geq F_2(s_m)$.

In summary, the optimal expected total system cost is increasing in the value of staleness function $F(s_m)$ when s_m is fixed.

C. Proof of Theorem 2

1) $-c_m(s_m, d_m)$ is nonincreasing in s_m for all $d_m \in D$.

Proof: When $d_m = 0$, in (8), we know $-c_m(s_m, d_m = 0) = -c_s(s_m)$. According to (11), $c_s(s_m)$ is an increasing function in system state s_m , therefore $-c_s(s_m, d_m)$ is nonincreasing when $d_m = 0$. When $d_m = 1$, $-c_m(s_m, d_m = 1) = -c_u$, where c_u is a constant. In summary, $-c_m(s_m, d_m)$ is nonincreasing in s_m for all $d_m \in D$.

2) $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}}$ is nondecreasing in s_m for all $s_{m+1} \in S$ and $d_m \in D$.

Proof: Let $s_m = a_m, s_{m+1} = a_n$. Due to the different cases for d_m , the proof process should be conducted, respectively.

When $d_m = 0$, according to (5), we have

$$\begin{aligned}
 \sum_{s_{m+1} \in S} P_{s_m, d_m=0, s_{m+1}} &= \sum_{s_{m+1} \in S} \frac{(\lambda_u)^{(s_{m+1}-s_m)} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(s_{m+1}-s_m+1)}} \\
 &= \sum_{s_{m+1}=a_n}^{\infty} \frac{(\lambda_u)^{(a_n-a_m)} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(a_n-a_m+1)}} \\
 &= \frac{\lambda_r}{(\lambda_u + \lambda_r)} \cdot \sum_{s_{m+1}=a_n}^{\infty} \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{(a_n-a_m)} \\
 &= \frac{\lambda_r}{(\lambda_u + \lambda_r)} \cdot \frac{\left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{(a_n-a_m)}}{1 - \frac{\lambda_u}{\lambda_u + \lambda_r}} \\
 &= \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{(a_n-a_m)}
 \end{aligned} \quad (A3)$$

Equation (A3) can be rewritten as $(\frac{\lambda_u + \lambda_r}{\lambda_u})^{(a_m - a_n)}$, where $\frac{\lambda_u + \lambda_r}{\lambda_u} > 1$ and $s_m = a_m$. Therefore, (A3) is nondecreasing in s_m when $d_m = 0$.

When $d_m = 1$, according to (7), we have

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s_m, d_m=1, s_{m+1}} &= \sum_{s_{m+1} \in S} \frac{(\lambda_u)^{s_{m+1}} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(s_{m+1}+1)}} \\ &= \sum_{s_{m+1}=a_n}^{\infty} \frac{(\lambda_u)^{a_n} \cdot \lambda_r}{(\lambda_u + \lambda_r)^{(a_n+1)}} = \frac{\lambda_r}{(\lambda_u + \lambda_r)} \\ &= \sum_{s_{m+1}=a_n}^{\infty} \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{a_n} = \frac{\lambda_r}{(\lambda_u + \lambda_r)} \cdot \frac{\left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{a_n}}{1 - \frac{\lambda_u}{\lambda_u + \lambda_r}} \\ &= \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{a_n} \end{aligned} \quad (\text{A4})$$

There is no a_m in (A4), therefore (A4) is a constant when $d_m = 1$.

Based on the above analysis, $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}}$ is nondecreasing in s_m for all $s_{m+1} \in S$ and $d_m \in D$.

3) $-c_m(s_m, d_m)$ is a superadditive function on $S \times D$.

Proof: Assume $s^+ \geq s^-$. First, when $d^+ > d^-$, i.e., $d^+ = 1$ and $d^- = 0$, according to Definition 2 and (8), we have

$$[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] = -c_u + [-c_s(s^-)]$$

and

$$[-c_m(s^+, d^-)] + [-c_m(s^-, d^+)] = -c_s(s^+) + (-c_u).$$

$c_s(s_m)$ is increasing in s_m , therefore $c_s(s^+) \geq c_s(s^-)$ for $s^+ \geq s^-$. Hence $[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] \geq [-c_m(s^+, d^-)] + [-c_m(s^-, d^+)]$ for $s^+ \geq s^-$ and $d^+ > d^-$, which means $-c_m(s_m, d_m)$ is a superadditive function.

When $d^+ = d^- = 0$, according to Definition 2 and (8), we have

$$[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] = [-c_s(s^+)] + [-c_s(s^-)]$$

and

$$[-c_m(s^+, d^-)] + [-c_m(s^-, d^+)] = [-c_s(s^+)] + [-c_s(s^-)]$$

Therefore, $[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] = [-c_m(s^+, d^-)] + [-c_m(s^-, d^+)]$ for $s^+ \geq s^-$ and $d^+ = d^- = 0$, which means $-c_m(s_m, d_m)$ is a superadditive function.

Last, when $d^+ = d^- = 1$, according to Definition 2 and (8), we have

$$[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] = (-c_u) + (-c_u)$$

and

$$[-c_m(s^+, d^-)] + [-c_m(s^-, d^+)] = (-c_u) + (-c_u).$$

Therefore, $[-c_m(s^+, d^+)] + [-c_m(s^-, d^-)] = [-c_m(s^+, d^-)] + [-c_m(s^-, d^+)]$ for $s^+ \geq s^-$ and $d^+ = d^- = 1$, which means $-c_m(s_m, d_m)$ is a superadditive function.

In summary, $-c_m(s_m, d_m)$ is a superadditive function on $S \times D$.

$\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot G_{m+1}^*$ is a superadditive function on $S \times D$ for all $s_{m+1} \in S$.

Proof: Assume $s^+ \geq s^-$ and $d^+ \geq d^-$. According to Definition 2, to prove that $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot G_{m+1}^*$ is a superadditive function, we must prove that

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s^+, d^+, s_{m+1}} \cdot G_{m+1}^* + \sum_{s_{m+1} \in S} P_{s^-, d^-, s_{m+1}} \cdot G_{m+1}^* \\ \geq \sum_{s_{m+1} \in S} P_{s^+, d^-, s_{m+1}} \cdot G_{m+1}^* + \sum_{s_{m+1} \in S} P_{s^-, d^+, s_{m+1}} \cdot G_{m+1}^* \end{aligned}$$

When $d^+ > d^-$, i.e., $d^+ = 1$ and $d^- = 0$, by Definition 2 and (A3) and (A4), we have

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s^+, d^+=1, s_{m+1}} + \sum_{s_{m+1} \in S} P_{s^-, d^-=0, s_{m+1}} \\ = \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{s_{m+1}} + \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{(s_{m+1} - s^-)} \end{aligned}$$

and

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s^+, d^-=0, s_{m+1}} + \sum_{s_{m+1} \in S} P_{s^-, d^+, s_{m+1}} \\ = 1, s_{m+1} = \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{(s_{m+1} - s^+)} + \left(\frac{\lambda_u}{\lambda_u + \lambda_r} \right)^{s_{m+1}} \end{aligned}$$

We have proved that $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}}$ is nondecreasing

in s_m , therefore for $s^+ \geq s^-$, we can get

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s^+, d^+=1, s_{m+1}} + \sum_{s_{m+1} \in S} P_{s^-, d^-=0, s_{m+1}} \\ \leq \sum_{s_{m+1} \in S} P_{s^+, d^-=0, s_{m+1}} + \sum_{s_{m+1} \in S} P_{s^-, d^+=1, s_{m+1}} \end{aligned} \quad (\text{A5})$$

Equation (A5) can be rewritten as

$$\begin{aligned} - \sum_{s_{m+1} \in S} (P_{s^+, d^+=1, s_{m+1}} + P_{s^-, d^-=0, s_{m+1}}) \\ \geq - \sum_{s_{m+1} \in S} (P_{s^+, d^-=0, s_{m+1}} + P_{s^-, d^+=1, s_{m+1}}) \end{aligned}$$

We know $V_m^* \geq 0$ for all $s_m \in S$, therefore we can get

$$\begin{aligned} - \sum_{s_{m+1} \in S} (P_{s^+, d^+=1, s_{m+1}} + P_{s^-, d^-=0, s_{m+1}}) V_{m+1}^* \\ \geq - \sum_{s_{m+1} \in S} (P_{s^+, d^-=0, s_{m+1}} + P_{s^-, d^+=1, s_{m+1}}) V_{m+1}^* \end{aligned} \quad (\text{A6})$$

Since $G_m^* = -V_m^*$, therefore (A6) can be rewritten as

$$\begin{aligned} \sum_{s_{m+1} \in S} P_{s^+, d^+, s_{m+1}} \cdot G_{m+1}^* + \sum_{s_{m+1} \in S} P_{s^-, d^-, s_{m+1}} \cdot G_{m+1}^* \\ \geq \sum_{s_{m+1} \in S} P_{s^+, d^-, s_{m+1}} \cdot G_{m+1}^* + \sum_{s_{m+1} \in S} P_{s^-, d^+, s_{m+1}} \cdot G_{m+1}^* \end{aligned}$$

Therefore, $\sum_{s_{m+1} \in S} P_{s_m, d_m, s_{m+1}} \cdot G_{m+1}^*$ is a superadditive function on $S \times D$ for all $s_{m+1} \in S$.

5) $-c_M(s_M, d_M)$ is nonincreasing in s_m .


```

Function getlm()
  m = M;
  lM* = ceil(cs(sM) ≥ cu);
  for m = M - 1 to 1
    if m = M - 1
      RSM0 = 0;
      v = 0;
      rsm = 0;
      tmp = 1;
      while tmp > 10e - 10 || rsm == 0
        if v < lM*
          rsm = cs(v);
        else
          rsm = cu;
        end
        tmp = Psm=v, dm=0, sm+1 · rsm;
        RSM0 = RSM0 + tmp;
        v = v + 1;
      end
    end
    RSM1 = RSM0 + cu;
    sm = 0;
    sn = sm + 1;
    while sn > 0
      RSM = 0;
      tmp = RSM0 - cs(sm);
      tmp = tmp - (λr/λu + λr) * Vm+1*(sm+1 = j);
      tmp = tmp * (λu + λr) / λu;
      tmp = tmp + cs(sm + 1);
      RSM = tmp;
      if RSM ≥ RSM1
        R = RSM1;
        JSM = [JSM, R];
        lm* = sm;
        break;
      else
        R = RSM;
        JSM = [JSM, R];
        R1 = RSM;
        sm = sm + 1;
        sn = sn + 1;
      end
    end
  end
end

```

Fig. A1. Algorithms for obtaining the control limits.

Proof: According to (8), when $m = M$ and $d_M = 0$, $-c_M(s_M, d_M = 0) = -c_s(s_M)$. According to the analysis before, $c_s(s_m)$ is an increasing function in system state s_m , therefore $-c_s(s_M, d_M)$ is nonincreasing when $d_M = 0$. When $d_M = 1$, $-c_M(s_M, d_M = 1) = -c_u$, where c_u is a constant. In summary, $-c_M(s_M, d_M)$ is nonincreasing in s_m for all $d_m \in D$.

REFERENCES

- [1] O. Zach, B. E. Munkvold, and D. H. Olsen, "ERP system implementation in SMEs: Exploring the influences of the SME context," *Enterprise Inf. Syst.*, vol. 8, no. 2, pp. 309–335, 2014.
- [2] M. G. Morris and V. Venkatesh, "Job characteristics and job satisfaction: Understanding the role of enterprise resource planning system implementation," *MIS Quart.*, vol. 34, no. 1, pp. 143–161, 2010.
- [3] Y. Zeng and M. J. Skibniewski, "Risk assessment for enterprise resource planning (ERP) system implementations: A fault tree analysis approach," *Enterprise Inf. Syst.*, vol. 7, no. 3, pp. 332–353, 2013.
- [4] Y.-Y. Huang and R. B. Handfield, "Measuring the benefits of ERP on supply management maturity model: A 'big data' method," *Int. J. Oper. Prod. Manage.*, vol. 35, no. 1, pp. 2–25, 2015.
- [5] R. Basu *et al.*, "An approach to identify issues affecting ERP implementation in Indian SMEs," *J. Ind. Eng. Manage.*, vol. 5, no. 1, pp. 133–154, 2012.
- [6] W.-H. Tsai *et al.*, "Investigation of the mediating effects of IT governance-value delivery on service quality and ERP performance," *Enterprise Inf. Syst.*, vol. 9, no. 2, pp. 139–160, 2015.
- [7] J. Manyika *et al.*, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, Chicago, 2011.
- [8] W. Zong *et al.*, "A discriminative and semantic feature selection method for text categorization," *Int. J. Prod. Econ.*, vol. 165, pp. 215–222, 2015.
- [9] J. Lu *et al.*, "A novel indexing method for improving timeliness of high-dimensional data," in *Proc. 20th Americas Conf. Inf. Syst.*, Savannah, GA, USA, 2014, pp. 1–12.
- [10] M. Bouzeghoub and V. Peralta, "A framework for analysis of data freshness," in *Proc. 2004 Int. Workshop Inf. Quality Inf. Syst.*, Paris, France, 2004, pp. 59–67.
- [11] X. Fang and R. Rachamadugu, "Policies for knowledge refreshing in databases," *Omega*, vol. 37, no. 1, pp. 16–28, 2009.
- [12] M. S. Islam, "Regulators of timeliness data quality dimension for changing data quality in information manufacturing system (IMS)," in *Proc. 3rd Int. Conf. Digit. Inf. Process. Commun.*, Dubai, United Arab Emirates, 2013, pp. 126–133.
- [13] B. Heinrich and M. Klier, "Metric-based data quality assessment—Developing and evaluating a probability-based currency metric," *Decis. Support Syst.*, vol. 72, pp. 82–96, 2015.
- [14] Y. Wand and R. Y. Wang, "Anchoring data quality dimensions in ontological foundations," *Commun. ACM*, vol. 39, no. 11, pp. 86–95, 1996.
- [15] Y. W. Lee and D. M. Strong, "Knowing-why about data processes and data quality," *J. Manage. Inf. Syst.*, vol. 20, no. 3, pp. 13–39, 2003.
- [16] D. Ballou *et al.*, "Modeling information manufacturing systems to determine information product quality," *Manage. Sci.*, vol. 44, no. 4, pp. 462–484, 1998.
- [17] I. C. Italiano and J. E. Ferreira, "Synchronization options for data warehouse designs," *Computer*, vol. 39, no. 3, pp. 53–57, 2006.
- [18] M. Mannino, S. N. Hong, and I. J. Choi, "Efficiency evaluation of data warehouse operations," *Decis. Support Syst.*, vol. 44, no. 4, pp. 883–898, 2008.
- [19] J. Wang *et al.*, "PL-Tree: An efficient indexing method for high-dimensional data," in *Proc. 13th Int. Symp. Adv. Spatial Temporal Databases*, Munich, Germany, 2013, pp. 183–200.
- [20] Q. Han and N. Venkatasubramanian, "Addressing timeliness/accuracy/cost tradeoffs in information collection for dynamic environments," in *Proc. 24th IEEE Int. Real-Time Syst. Symp.*, Cancun, Mexico, 2003, pp. 108–117.
- [21] C. Chen and J. Chen, "Timeliness evaluation of task-oriented networked space-based information system," *J. Syst. Eng. Electron.*, vol. 22, no. 4, pp. 621–627, 2011.
- [22] K. M. Chandy, J. C. Browne, C. W. Dissly, and W. R. Uhrig, "Analytic models for rollback and recovery strategies in data base systems," *IEEE Trans. Softw. Eng.*, vol. SE-1, no. 1, pp. 100–110, Mar. 1975.
- [23] J. Srivastava and D. Rotem, "Analytical modeling of materialized view maintenance," in *Proc. 7th ACM SIGACT-SIGMOD-SIGART Symp. Principles Database Syst.*, Austin, TX, USA, 1988, pp. 126–134.
- [24] A. Segev and W. P. Fang, "Optimal update policies for distributed materialized views," *Manage. Sci.*, vol. 37, no. 7, pp. 851–870, Jul. 1991.
- [25] B. Adelberg, H. Garcia-Molina, and B. Kao, "Applying update streams in a soft real-time database system," in *Proc. 1995 ACM SIGMOD Int. Conf. Manage. Data*, San Jose, CA, USA, 1995, pp. 245–256.

- [26] Y. B. Ling and J. Mi, "An optimal trade-off between content freshness and refresh cost," *J. Appl. Probab.*, vol. 41, no. 3, pp. 721–734, Sep. 2004.
- [27] D. Dey, Z. Zhang, and P. De, "Optimal synchronization policies for data warehouses," *INFORMS J. Comput.*, vol. 18, no. 2, pp. 229–242, 2006.
- [28] M. V. Mannino and Z. Walter, "A framework for data warehouse refresh policies," *Decis. Support Syst.*, vol. 42, no. 1, pp. 121–143, 2006.
- [29] X. Fang, O. R. L. Sheng, and P. Goes, "When is the right time to refresh knowledge discovered from data?" *Oper. Res.*, vol. 61, no. 1, pp. 32–44, 2013.
- [30] V. Paxson and S. Floyd, "Wide area traffic—The failure of poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.
- [31] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2009.



Wei Zong received the Ph.D. degree in management science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2016.

She is currently a Faculty Member with the School of Economics and Management, Xidian University. Her research articles have been published in the *International Journal of Production Economics* and *Enterprise Information Systems*. Her current research interests include data quality management, data fusion, and data mining.



Feng Wu received the Ph.D. degree in management science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2003.

He is currently a Professor with the Department of Industrial Engineering, Xi'an Jiaotong University. He has authored or co-authored articles published in refereed journals such as the *International Journal of Production Research*, *Annals of Operations Research*, the *European Journal of Operational Research*, the *International Journal of Production Economics*, *Enterprise Information Systems*, and *Computers and Industrial Engineering*. In addition, he has also published more than 50 conference papers. His current research interests include outsourcing decisions and control, procurement risk management, data quality, supply chain management, and intelligent manufacturing management.



Zhengrui Jiang received the Ph.D. degree in management science from the University of Texas at Dallas, Dallas, TX, USA.

He is the Thome Professor in Business and an Associate Professor of information systems with the College of Business, Iowa State University, IA, USA. He has published in leading academic journals including *Information Systems Research*, *Management Science*, *MIS Quarterly*, the *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, and the *Journal of Management Information Systems*. His primary research interests include business intelligence/analytics, diffusion of innovations, and economics of information goods.

He currently serves as an Associate Editor for *MIS Quarterly*. He also served as a Program Co-Chair for the 2014 Midwest Association of Information Systems Conference and the 2015 Big XII+ MIS Research Symposium.