Storage Space Allocation Strategy for Digital Data with Message Importance

Shanyun Liu [1,2], **Rui She** [1,2], **Zheqi Zhu** [1,2] and **Pingyi Fan** [1,2,*]

[1]   Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
     liushany16@mails.tsinghua.edu.cn (S.L.); sher15@mails.tsinghua.edu.cn (R.S.);
     zhuzq18@mails.tsinghua.edu.cn (Z.Z.)
[2]   Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China
[*]   Correspondence: fpy@tsinghua.edu.cn; Tel.: +86-010-6279-6973

Abstract: This paper mainly focuses on the problem of lossy compression storage based on the data value that represents the subjective assessment of users when the storage size is still not enough after the conventional lossless data compression. To this end, we transform this problem to an optimization, which pursues the least importance-weighted reconstruction error in data reconstruction within limited total storage size, where the importance is adopted to characterize the data value from the viewpoint of users. Based on it, this paper puts forward an optimal allocation strategy in the storage of digital data by the exponential distortion measurement, which can make rational use of all the storage space. In fact, the theoretical results show that it is a kind of restrictive water-filling. It also characterizes the trade-off between the relative weighted reconstruction error and the available storage size. Consequently, if a relatively small part of total data value is allowed to lose, this strategy will improve the performance of data compression. Furthermore, this paper also presents that both the users' preferences and the special characteristics of data distribution can trigger the small-probability event scenarios where only a fraction of data can cover the vast majority of users' interests. Whether it is for one of the reasons above, the data with highly clustered message importance is beneficial to compression storage. In contrast, from the perspective of optimal storage space allocation based on data value, the data with a uniform information distribution is incompressible, which is consistent with that in the information theory.

Keywords: lossy compression storage; optimal allocation strategy; weighted reconstruction error; message importance measure; importance coefficient

1. Introduction

As large amounts of mobile devices such as Internet of things (IoT) devices or smartphones are utilized, the contradiction between limited storage space and sharply increasing data deluge becomes increasingly serious in the era of big data [1,2]. This exceedingly massive data makes the conventional data storage mechanisms inadequate within a tolerable time, and therefore the data storage is one of the major challenges in big data [3]. Note that storing all the data becomes more and more dispensable nowadays, and it is also not conducive to reduce data transmission costs [4,5]. In fact, data compression storage is widely adopted in many applications, such as IoT [2], industrial data platform [6], bioinformatics [7], wireless networking [8]. Thus, the research on data compression storage becomes increasingly paramount and compelling nowadays.

In conventional source coding, data compression is carried out by removing the data redundancy, where short descriptions are assigned to the most frequent class [9]. Based on it, the tight bounds for lossless data compression are given. In order to further increase the compression rate, one needs to use

more information. A quintessential example is to use some side information [10]. Another possible solution is to compress the data with quite a few losses first and then reconstruct them with acceptable distortion, which is referred to as lossy compression [11–13]. Some adaptive compressions are adopted extensively. For example, Reference [14] proposed an adaptive compression scheme in IoT systems, and Reference [15] investigated the backlog-adaptive source coding system in terms of age of information. In fact, most of the previous compression methods usually carried out compression by means of contextual data or leveraging data transformation techniques [4].

Although these previous methods of data compression perform satisfactorily in their respective application scenarios, there is still much room for improvement when facing rapidly growing large-scale data. Moreover, they also do not take the data value into account. This paper focuses on the problem of how to further compress data with acceptable distortion to implement the specified requirements in data storage when the storage size is still not enough to guarantee the lossless storage after the conventional lossless data compression. This paper will realize this goal by reallocating storage space based on the data value which represents the subjective assessment of users. Here, we take the importance-aware weighting in the weighted reconstruction error to measure the total cost in data storage with unequal costs.

Generally, users prefer to care about the crucial part of data that attracts their attention rather than the whole data itself. In many real-world applications, such as cost-sensitive learning [16–18] and unequal error protection [19,20], different errors bring different costs. To be specific, the distortion in the data that users care about may be catastrophic if the loss of some data being insignificant for users is allowed. Similar to coresets [21], the data needing to be processed was reduced to those users as the main focus rather than the whole data set. Unlike coresets, the data needing to be processed in this paper no longer pursues approximately representing the raw data, and it is expected to minimize the storage cost with respect to the importance weighting value. In fact, although the data deluge sharply increases, the significant data that users care about is still rare in a lot of scenarios of big data. In this sense, it can be regarded as the sparse representation from the perspective of the data value, and we can use it to compress data.

Alternatively, it is interesting to achieve data compression by storing a fraction of data, which preserves as much information as possible regarding the data that users care about [22,23]. This paper also employs this strategy. However, there are subtle but critical differences between the compression storage strategy proposed in this paper with those in Reference [22,23]. In fact, Reference [22] focused on Pareto-optimal data compression, which presents the trade-off between retained entropy and class information. However, this paper puts forward an optimal compression storage strategy for digital data from the viewpoint of message importance, and it gives the trade-off between the relative weighted reconstruction error (RWRE) and the available storage size. Furthermore, the compression method based on message importance was preliminarily discussed in Reference [23] to solve the big data storage problem in wireless communications, while this paper will aim to discuss the optimal storage space allocation strategy with limited storage space, in general, based on message importance. Moreover, the constraints are also different. That is, the available storage size is limited in this paper, while the total code length of all the events is given in Reference [23].

From users' attention viewpoint, the data value can be considered as the subjective assessment of users on the importance of data. Actually, much of the research in the last decade suggested that the study from the perspective of message importance is rewarding to obtain new findings [20,24,25]. Thus, there may be effective performance improvement in storage systems when taking message importance into account. For example, Reference [26] discussed the lossy image compression method with the aid of a content-weighted importance map. Since any quantity can be seen as important if it agrees with the intuitive characterization of the user's subjective degree of concern of data, the cost in data reconstruction for specific user preferences is regarded as the importance in this paper, which will be used as the weight in the weighted reconstruction error.

Since we desire to achieve data compression by keeping only a small portion of important data and abandoning less important data, this paper mainly focuses on the case where only a fraction of data take up the vast majority of the users' interests. Actually, these scenarios are not rare in big data. A quintessential example should be cited that the minority subset detection is overwhelmingly paramount in intrusion detection [27,28]. Moreover, this phenomenon is also exceedingly typical in financial crime detection systems for the fact that only a few illicit identities catch our eyes to prevent financial frauds [29]. Actually, when a certain degree of information loss can be acceptable, people prefer to take high-probability events for granted and abandon them to maximize the compressibility. These cases are referred to as *small-probability event scenarios* in this paper. In order to depict the message importance in small-probability event scenarios, message importance measure (MIM) was proposed in Reference [30]. Furthermore, MIM is fairly effective in many applications of big data, such as IoT [31], mobile edge computing [32]. In addition, Reference [33] expanded MIM to the general case, and it presented that MIM can be adopted as a special weight in designing the recommendation system. Since there is no universal data value model, we might as well take the case where the MIM describes the cost of the error as a quintessential example to analyze the property of the optimal storage space allocation strategy.

In this paper, we firstly propose a particular storage space allocation strategy for digital data on the best effort in minimizing the importance-weighted reconstruction error when the total available storage size is provided. For digital data, we formulate this problem as an optimization problem, and present the optimal storage strategy by means of a kind of restrictive water-filling. For the given available storage size, the storage size is mainly determined by the values of message importance and probability distribution of event class in a data sequence. In fact, this optimal allocation strategy adaptively prefers to provide more storage size for crucial data classes in order to make the rational use of resources, which is in accord with the cognitive mechanism of human beings.

Afterward, we focus on the properties of this optimal storage space allocation strategy when the importance weights are characterized by MIM. It is noted that there is a trade-off between the RWRE and the available storage size. The constraints on the performance of this storage system are true, and they depend on the importance coefficient and the probability distribution of events classes. On the one hand, the RWRE increases with the increasing of the absolute value of importance coefficient for the fact that the overwhelming majority of important information will gather in a fraction of data as the importance coefficient increases to negative/positive infinity, which suggests the influence of users' preferences. On the other hand, the compression performance is also affected by probability distribution of event classes. In fact, the more closely the probability distribution matches the requirement of the small-probability event scenarios, the more effective this compression strategy becomes. Furthermore, it is also obtained that the RWRE in a uniform distribution is larger than any other distributions for the same available storage size. In this regard, the uniform distribution is incompressible from the perspective of optimal storage space allocation based on data value, which is consistent with the conclusion in information theory [34].

The main contributions of this paper can be summarized as follows. (1) It proposes a new digital data compression strategy taking message importance into account, which can help improve the design of a big data storage system. (2) We illuminate the properties of this new method, which can characterize the trade-off between the RWRE and the available storage size. (3) It shows that the data with highly clustered message importance is beneficial to compression storage, and it also finds that the data with a uniform information distribution is incompressible from the perspective of optimal storage space allocation based on data value, which is consistent with that in information theory.

The rest of this paper is organized as follows. The system model is introduced in Section 2, including the definition of weighted reconstruction error, distortion measure, and problem formulation. In Section 3, we solve the problem of optimal storage space allocation in three kinds of system models and give the solutions. The properties of this optimal storage space allocation strategy based on MIM are fully discussed in Section 4. The effects of the importance coefficient and the probability of event

classes on RWRE are also investigated in detail. Section 5 illuminates the properties of this optimal storage strategy when the importance weight is characterized by Non-parametric MIM. The numerical results are shown and discussed in Section 6, which verifies the validity of the developed theoretical results in this paper. Finally, we give the conclusion in Section 7.

2. System Model

This section introduces the system model, including the definition of the weighted reconstruction error, the modeling of distortion measure, in order to illustrate how we formulate the lossy compression problem as an optimization problem for digital data based on message importance. In order to make the formulation and discussion more clear, the main notations in this paper are listed in Table 1.

Table 1. Notations.

Notation	Description
$x = x_1, x_2, \ldots, x_k, \ldots, x_K$	The sequence of raw data
$\hat{x} = \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, \ldots, \hat{x}_K$	The sequence of compressed data
S_x	The storage size of x
$D_f(S_{x1}, S_{x2})$	The distortion measure function between S_{x1} and S_{x2} in data reconstruction
n	The number of event classes
$\{a_1, a_2, \ldots, a_n\}$	The alphabet of raw data
$\{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n\}$	The alphabet of compressed data
$W = \{W_1, W_2, \ldots, W_n\}$	The error cost for the reconstructed data
$P = \{p_1, p_2, \ldots, p_n\}$	The probability distribution of data class
$D(x, W)$	The weighted reconstruction error
$D_r(x, W), D_r(W, L, l)$	The relative weighted reconstruction error
$L = L_1, L_2, \ldots, L_n$	The storage size of raw data
$l = l_1, l_2, \ldots, l_n$	The storage size of compressed data
l_i^*	The round optimal storage size of the data belonging to the i-th class
T	The maximum available average storage size
ϖ	The importance coefficient
γ_p	$\gamma_p = \sum_{i=1}^n p_i^2$
α_1, α_2	$\alpha_1 = \arg\min_i p_i$ and $\alpha_2 = \arg\max_i p_i$
$L(\varpi, p)$	The message importance measure, which is given by $L(\varpi, p) = \ln \sum_{i=1}^n p_i e^{\varpi(1-p_i)}$
Δ	The average compressed storage size of each data, which is given by $\Delta = L - T$
$\Delta^*(\delta)$	The maximum available Δ for the given supremum of the RWRE δ
$\mathcal{L}(P)$	The non-parametric message importance measure, which is given by $\mathcal{L}(P) = \ln \sum_{i=1}^n p_i e^{(1-p_i)/p_i}$

2.1. Modeling Weighted Reconstruction Error Based on Message Importance

The data storage system may lack storage space frequently when facing a super-large scale of data to store. When the storage size is still not enough after the lossless conventional data compression, the optimum allocation of storage space based on data value may be imperative. For this purpose, we consider the following storage system, which stores K pieces of data. Let $x = x_1, x_2, \ldots, x_k, \ldots, x_K$ be the sequence of raw data. Assume that all the data redundancy have been removed after the lossless conventional data compression, and each data x_k needs to take up storage space with size of Sx_k if this data can be recovered without any distortion. However, in many scenarios of big data, the storage size is still not enough in this case. That is to say, the actual required storage space $\sum_{k=1}^K Sx_k$ is larger than the maximum available storage space TK, where T is the maximum available average storage size.

In fact, users prefer to care about the paramount part of data that attracts their attention rather than the whole data itself. In this perspective, storing all data without distortion may be unnecessary. Considering that the natural distribution of storage space is not invariably reasonable and the high value data in big data is usually sparse, the rational storage space allocation by minimizing the loss of data value may solve the above problem of insufficient storage space, if a certain amount of data value is allowed to be lost. After the data compression by means of the rational storage space allocation, we use $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, \ldots, \hat{x}_K$ to denote the compressed data sequence, and assume that the compressed data $\hat{x}_k$ takes up storage space with size of $S\hat{x}_k$ in practice for $1 \leq k \leq K$.

The lossy data compression usually pursues the least the storage cost while retaining as much information users required as possible [22]. In the lossless conventional data compression, the costs of different data are assumed to be the same. However, different kinds of errors may result in unequal costs in many real-world applications [16–19]. In this model, we use the notation W_k to denote the error cost for the reconstructed data. Namely, W_k is with respect to the data value of data x_k, and it is regarded as the message importance in this paper. Here, we define the weighted reconstruction error to describe the total cost in data storage with unequal costs, which is given by

$$D(\boldsymbol{x}, \boldsymbol{W}) = \frac{1}{K} \sum_{k=1}^{K} W_k D_f(Sx_k, S\hat{x}_k), \tag{1}$$

where $D_f(Sx_k, S\hat{x}_k)$ characterizes the distortion between the raw data and the compressed data in data reconstruction, which characterizes the loss degree of data value with allocated storage size.

Consider the situation where the data is stored according to its category for easier retrieval, which can also make the recommendation system based on it more effective [33]. Since data classification is becoming increasingly convenient and accurate nowadays due to the rapid development of machine learning [35,36], this paper assumes that the event class can be easily detected and known in the storage system. Moreover, assume the data that belongs to the same class has the same importance-weight and occupies the same storage size. Hence, $\boldsymbol{x}$ can be seen as a sequence of K symbols from an alphabet $\{a_1, a_2, \ldots, a_n\}$ where a_i represents event class i. This storage model is summarized and shown in Figure 1. In this case, the weighted reconstruction error based on importance is formulated as

$$D(\boldsymbol{x}, \boldsymbol{W}) = \sum_{i=1}^{n} \frac{N(a_i|\boldsymbol{x})}{K} W_i D_f(Sa_i, S\hat{a}_i) \tag{2}$$

$$= \sum_{i=1}^{n} p_i W_i D_f(Sa_i, S\hat{a}_i), \tag{2a}$$

where $N(a_i|\boldsymbol{x})$ is the number of times the i-class occurs in the sequence $\boldsymbol{x}$. Let $p_i = N(a_i|\boldsymbol{x})/K$ denote the probability of event class i in data sequence $\boldsymbol{x}$.
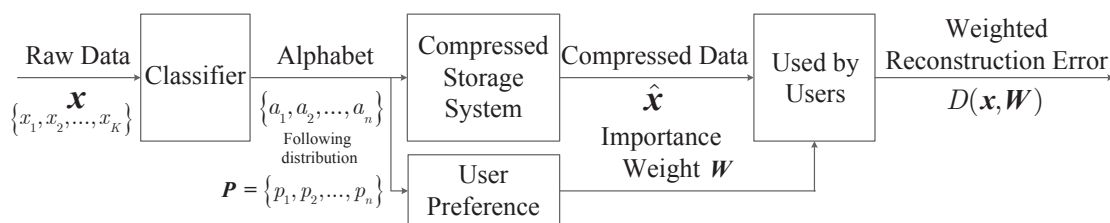

Figure 1. Pictorial representation of the system model.

2.2. Modeling Distortion between the Raw Data and the Compressed Data

We focus on the formula of D_f in this part, which characterizes the distortion between the raw data and the compressed data with specified storage size. Usually, there is no universal characterization of distortion measure, especially in speech coding and image coding [34]. In fact, D_f should characterize the loss degree of data value with allocated storage size. In this respect, the conventional distortion measures are not appropriate since they do not take unequal costs into account. In order to facilitate the analysis and design, this paper proposes an exponential distortion measure to discuss the following special case.

We assume that the data is digital and ignore the storage formats and standards in concrete application environments. On its application fields, it may be useful in some scenarios with counting systems, such as finance, or medicine, as the general merchandise. Let the description of the raw data a_i be L_i bits, and $a_i = \sum_{j=0}^{L_i-1} b_j \times r^j$ where r is radix ($r > 1$). Actually, the radix represents the base of the system in practical application, such as $r = 2$ in a binary system. In particular, L_i will approach

the infinite number if a_i is an arbitrary real number. When the storage size is still not enough after the lossless conventional data compression, there is only l_i bits assigned to it in order to compress data further based on the message importance. For convenience, the smaller $(L_i - l_i)$ numbers are discarded in this process. When restoring the compressed data, the discarded digits are set to the same pre-specified number or random numbers in the actual system. Let b_j^* be the $(j + 1)$-th discarded digit for $j = 0, 1, \cdots, L_i - l_i - 1$, and assume that b_j^* is a random number in $\{0, \ldots, r - 1\}$. In this case, the compressed data is $\hat{a}_i = \sum_{j=L_i-l_i}^{L_i-1} b_j \times r^j + \sum_{j=0}^{L_i-l_i-1} b_j^* \times r^j$. As a result, the absolute error is $|a_i - \hat{a}_i|$, which meets

$$|a_i - \hat{a}_i| = |\sum_{j=0}^{L_i-l_i-1} (b_j - b_j^*) \times r^j| \le r^{L_i-l_i} - 1. \tag{3}$$

When $l_i = 0$, which means there is no information stored, the supremum of absolute error reaches the maximum and it is $|a_i - \hat{a}_i| \le r^{L_i} - 1$. In order to better weigh the different costs, we define the relative error by normalizing the absolute error to the interval $[0, 1]$ based on the above maximum absolute error $r^{L_i} - 1$. Moreover, we adopt the supremum of this relative error as the distortion measure D_f, which is given by

$$D_f(Sa_i, S\hat{a}_i) = D_f(L_i, l_i) = \frac{r^{L_i-l_i} - 1}{r^{L_i} - 1}. \tag{4}$$

In particular, we obtain $D_f(L_i, L_i) = 0$ and $D_f(L_i, 0) = 1$. Moreover, it is easy to check that $0 \le D_f(L_i, l_i) \le 1$ and $D_f(L_i, l_i)$ decreases with the increasing of l_i. In fact, D_f can be regarded as the percentage of data value loss in this case. Thus, the weighted reconstruction error in Equation (1) represents the total cost in data storage based on the loss degree.

In this stored procedure, the compression rate is $(\sum_{i=1}^n p_i l_i)/(\sum_{i=1}^n p_i L_i)$, and the total saving storage size is $\sum_{i=1}^n p_i(L_i - l_i)K$. Actually, K denotes the number of data, and it is extremely big due to the sharply increasing data deluge in the era of big data. Therefore, although $(L_i - l_i)$ is not always large, the saving storage size is still exceedingly substantial since K is exceedingly large.

Furthermore, to simplify the comparisons under different conditions, the weighted reconstruction error is also normalized to the relative weighted reconstruction error (RWRE). In fact, the RWRE characterizes the relative total cost in the data compression, and it is given by

$$D_r(\boldsymbol{x}, \boldsymbol{W}) = D_r(\boldsymbol{W}, \boldsymbol{L}, \boldsymbol{l}) = \frac{D(\boldsymbol{x}, \boldsymbol{W})}{\max_{l_i} D(\boldsymbol{x}, \boldsymbol{W})} = \frac{\sum_{i=1}^n p_i W_i D_f(L_i, l_i)}{\sum_{i=1}^n p_i W_i} = \frac{\sum_{i=1}^n p_i W_i \frac{r^{L_i-l_i}-1}{r^{L_i}-1}}{\sum_{i=1}^n p_i W_i}, \tag{5}$$

where $\boldsymbol{L} = \{L_1, \ldots, L_n\}$ and $\boldsymbol{l} = \{l_1, \ldots, l_n\}$.

2.3. Problem Formulation

2.3.1. General Storage System

In fact, the actual storage size of each data after the compression can then be expressed as $\sum_{i=1}^n p_i l_i$. For each given maximum available storage space constraint $\sum_{i=1}^n p_i l_i \le T$, where T denotes the maximum available average storage size, we shall optimize the storage resources allocation strategy of this system by minimizing the RWRE, which can be expressed as

$$\mathcal{P}_1: \min_{l_i} D_r(\boldsymbol{x}, \boldsymbol{W}) \tag{6}$$

$$\text{s.t.} \sum_{i=1}^n p_i l_i \le T \tag{6a}$$

$$0 \le l_i \le L_i \text{ for } i = 1, 2, \ldots, n. \tag{6b}$$

The storage systems, which can be characterized by Problem $\mathcal{P}_1$, are referred to as the *general storage system*.

Remark 1. *In fact, this paper focuses on allocating resources by category with taking message importance into account, while the conventional source coding searches the shortest average description length of a random variable.*

2.3.2. Ideal Storage System

In practice, the storage size of raw data is usually assigned to be the same for ease of use. Thus, we mainly consider the case where the original storage size of each data is the same, and use L to denote it (i.e., $L_i = L$ for $i = 1, 2, \ldots, n$). As a result, we have

$$\min_{l_i} D_r(x, W) = \frac{r^L \min\limits_{l_i} \sum\limits_{i=1}^{n} p_i W_i r^{-l_i}}{(r^L - 1) \sum\limits_{i=1}^{n} p_i W_i} - \frac{1}{r^L - 1}. \tag{7}$$

Thus, the problem $\mathcal{P}_1$ can be rewritten as

$$\mathcal{P}_2 : \min_{l_i} \ \sum_{i=1}^{n} p_i W_i r^{-l_i} \tag{8}$$

$$\text{s.t.} \ \sum_{i=1}^{n} p_i l_i \leq T \tag{8a}$$

$$0 \leq l_i \leq L \text{ for } i = 1, 2, \ldots, n. \tag{8b}$$

For convenience, we use the *ideal storage system* to represent the storage systems, which can be described by Problem $\mathcal{P}_2$. Moreover, we will mainly focus on the characteristics of the solutions in Problem $\mathcal{P}_2$ in this paper.

2.3.3. Quantification Storage System

A *quantification storage system* quantizes and stores the real data acquired from sensors in the real world. The data is usually a real number, which requires an infinite number of bits to describe it accurately. That is, the original storage size of each class approaches the infinite number, (i.e., $L_i = L \to +\infty$ for $i = 1, 2, \ldots, n$), in this case. As a result, the RWRE can be rewritten as

$$D_r(x, W) = \lim_{L \to \infty} \left\{ \frac{\sum\limits_{i=1}^{n} p_i W_i r^{-l_i}}{(1 - r^{-L}) \sum\limits_{i=1}^{n} p_i W_i} - \frac{1}{r^L - 1} \right\} = \frac{\sum\limits_{i=1}^{n} p_i W_i r^{-l_i}}{\sum\limits_{i=1}^{n} p_i W_i}. \tag{9}$$

Therefore, the problem $\mathcal{P}_1$ in this case is reduced to

$$\mathcal{P}_3 : \min_{l_i} \ \sum_{i=1}^{n} p_i W_i r^{-l_i} \tag{10}$$

$$\text{s.t.} \ \sum_{i=1}^{n} p_i l_i \leq T \tag{10a}$$

$$l_i \geq 0 \text{ for } i = 1, 2, \ldots, n. \tag{10b}$$

3. Optimal Allocation Strategy with Limited Storage Space

In this section, we shall first solve the problem $\mathcal{P}_1$ and give the solutions. In fact, the solutions provide the optimal storage space allocation strategy for digital data on the best effort in minimizing the relative weighted reconstruction error (RWRE) when the total available storage size is limited.

Then, the problem $\mathcal{P}_2$ will be solved, the solutions of which characterize the optimal storage space allocation strategy with the same original storage size. Moreover, we shall also discuss the solutions in the case where the original storage size of each class approaches the infinite number by studying the problem $\mathcal{P}_3$.

3.1. Optimal Allocation Strategy in General Storage System

Theorem 1. *For a storage system with probability distribution $(p_1, p_2, \ldots, p_n)$, L_i is the storage size of the raw data of the class i for $i = 1, 2, \ldots, n$. For a given maximum available average storage size T $(0 \leq T \leq \sum_{i=1}^{n} p_i L_i)$, when the radix is r $(r > 1)$, the solution of Problem $\mathcal{P}_1$ is given by*

$$
l_i = \begin{cases} 0 & \text{if } l_i < 0, \\ \dfrac{\ln(\ln r) + \ln W_i - \ln(1 - r^{-L_i}) - \ln \lambda^*}{\ln r} & \text{if } 0 \leq l_i \leq L_i, \\ L_i & \text{if } l_i > L_i, \end{cases} \tag{11}
$$

where λ^* is chosen so that $\sum_{i=1}^{n} p_i l_i = T$.

Proof. By means of Lagrange multipliers and Karush–Kuhn–Tucher conditions, when ignoring the constant $\sum_{i=1}^{n} p_i W_i$, we set up the functional

$$
J = \sum_{i=1}^{n} p_i W_i \frac{r^{L_i - l_i} - 1}{r^{L_i} - 1} + \lambda^* \left(\sum_{i=1}^{n} p_i l_i - T \right) + \mu_1 (l_1 - L_1) + \cdots + \mu_n (l_n - L_n). \tag{12}
$$

Differentiating with respect to l_i and setting the derivative to zero, we have

$$
\frac{\partial J}{\partial l_i} = -p_i W_i \frac{r^{-l_i}}{1 - r^{-L_i}} + \lambda^* p_i + \mu_i = 0 \text{ for } i = 1, 2, \ldots, n \tag{13}
$$

$$
\sum_{i=1}^{n} p_i l_i - T = 0 \tag{13a}
$$

$$
\mu_i (l_i - L_i) = 0 \text{ for } i = 1, 2, \ldots, n \tag{13b}
$$

$$
l_i - L_i \leq 0 \text{ for } i = 1, 2, \ldots, n \tag{13c}
$$

$$
\mu_i \geq 0 \text{ for } i = 1, 2, \ldots, n \tag{13d}
$$

$$
l_i \geq 0 \text{ for } i = 1, 2, \ldots, n \tag{13e}
$$

Hence, we obtain

$$
l_i = \frac{\ln p_i + \ln(\ln r) + \ln W_i - \ln(1 - r^{-L_i}) - \ln(\lambda^* p_i + \mu_i)}{\ln r}. \tag{14}
$$

First, it is easy to check that Equations (13b)–(13d) hold when $\mu_i = 0$ and $l_i \leq L_i$. Hence, we have

$$
l_i = \frac{\ln(\ln r) + \ln W_i - \ln(1 - r^{-L_i}) - \ln \lambda^*}{\ln r}. \tag{15}
$$

Second, if l_i in Equation (14) is larger than L_i, we will have $\mu_i > 0$ and $l_i = L_i$ due to Equations (13b)–(13d).

Third, if $l_i < 0$, we will let $l_i = 0$ according to Equation (13e).

Moreover, λ^* is chosen so that $\sum_{i=1}^{n} p_i l_i = T$ due to Equation (13a).

Therefore, based on the discussion above, we get Equation (11) in order to ensure $0 \leq l_i \leq L_i$. $\square$

Remark 2. *Let $\tilde{N}$ be the number of l_i which meets $0 \leq l_i \leq L_i$ and $\{I_j, j = 1, 2, \ldots, \tilde{N}\}$ is part of the sequence of $\{1, 2, \ldots, N\}$ which satisfies $0 \leq \ln(\ln r) + \ln W_{I_j} - \ln(1 - r^{-L_{I_j}}) - \ln \lambda^* \leq L_{I_j} \ln r$.*

Furthermore, $\{T_j, j = 1, 2, \ldots, \tilde{N}_L\}$ is used to denote the part of the sequence of $\{1, 2, \ldots, N\}$ which satisfies $\ln(\ln r) + \ln W_{T_j} - \ln(1 - r^{-L_{T_j}}) - \ln \lambda^* > L_{T_j} \ln r$.

Substituting Equation (11) in the constraint $\sum_{i=1}^{n} p_i l_i = T$, we have

$$\ln \lambda^* = \ln \ln r + \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j} - \sum_{j=1}^{\tilde{N}} p_{I_j} \ln(1 - r^{-L_{I_j}}) - \ln r(T - \sum_{j=1}^{\tilde{N}_L} p_{T_j} L_{T_j})}{\sum_{j=1}^{\tilde{N}} p_{I_j}}. \tag{16}$$

Hence, for $0 \le l_i \le L$, we obtain

$$l_i = \frac{T - \sum_{j=1}^{\tilde{N}_L} p_{T_j} L_{T_j}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\ln W_i}{\ln r} - \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j}}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln(1 - r^{-L_{I_j}})}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}}. \tag{17}$$

In fact, T, p_i, r, L_i are usually constraints for a given storage system, and therefore l_i is only determined by the second and the third items on the right side of Equation (17), which means the storage size depends on the message importance and the probability distribution of class for the given available storage size.

Remark 3. *Since the actual compressed storage size l_i^* must be an integer, the actual storage size allocation strategy is*

$$l_i^* = \min\left(\left\lfloor \frac{T - \sum_{j=1}^{\tilde{N}_L} p_{T_j} L_{T_j}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\ln W_i}{\ln r} - \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j}}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln(1 - r^{-L_{I_j}})}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} \right\rfloor^+, L_i\right), \tag{18}$$

where $(x)^+$ is equal to x when $x \ge 0$, and it is zero when $x < 0$. In addition, $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.

3.2. Optimal Allocation Strategy in Ideal Storage System

Then, we pay attention to the case where the original storage size of each data is the same. Based on Theorem 1, we get the following corollary in the ideal storage system.

Corollary 1. *For a storage system with probability distribution $(p_1, p_2, \ldots, p_n)$, the original storage size of each class is the same, which is given by $L_i = L$ for $i = 1, 2, \ldots, n$. For a given maximum available average storage size T $(0 \le T \le L)$, when the radix is r $(r > 1)$, the solution of Problem $\mathcal{P}_2$ is given by*

$$l_i = \begin{cases} 0 & \text{if } l_i < 0, \\ \dfrac{\ln(\ln r) + \ln W_i - \ln \lambda}{\ln r} & \text{if } 0 \le l_i \le L, \\ L & \text{if } l_i > L, \end{cases} \tag{19}$$

where λ is chosen so that $\sum_{i=1}^{n} p_i l_i = T$.

Proof. Let $\lambda = \lambda^*(1 - r^{-L})$ and $L_i = L$ for $i = 1, 2, \ldots, n$. Substituting them in Equation (11), we find that l_i in this case can be rewritten as Equation (19). $\square$

Substituting Equation (19) in the constraint $\sum_{i=1}^{n} p_i l_i = T$, we obtain

$$\ln \lambda = \ln \ln r + \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j} - \ln r(T - T_{N_L})}{\sum_{j=1}^{\tilde{N}} p_{I_j}}, \tag{20}$$

where $\tilde{N}$, $\tilde{N}_L$, I_j, T_j is still given by Remark 2 with letting $\lambda = \lambda^*(1 - r^{-L})$. In addition, $T_{N_L} = \sum_{j=1}^{\tilde{N}_L} p_{T_j} L$. Hence, for $0 \leq l_i \leq L$, we obtain

$$l_i = \frac{T - T_{N_L}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\ln W_i}{\ln r} - \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j}}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}}. \tag{21}$$

Remark 4. *Since the actual compressed storage size l_i^* must be an integer, the actual storage size allocation strategy is*

$$l_i^* = \min\left(\left\lfloor \frac{T - T_{N_L}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{\ln W_i}{\ln r} - \frac{\sum_{j=1}^{\tilde{N}} p_{I_j} \ln W_{I_j}}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} \right\rfloor^+, L\right). \tag{22}$$

Remark 5. *When $\tilde{N} = n$, $0 \leq l_i \leq L$ always holds for $1 \leq i \leq n$, and the actual storage size is given by*

$$l_i^* = \left\lfloor T + \frac{\ln W_i - \sum_{i=1}^{n} p_i \ln W_i}{\ln r} \right\rfloor. \tag{23}$$

In order to illustrate the geometric interpretation of this algorithm, let

$$\beta = \frac{\ln \ln r - \ln \lambda}{\ln r}. \tag{24}$$

Hence, the optimal storage size can be simplified to

$$l_i = \begin{cases} 0, & \text{if } \beta - \dfrac{\ln(1/W_i)}{\ln r} < 0. \\[2mm] \beta - \dfrac{\ln(1/W_i)}{\ln r} & \text{if } 0 \leq \beta - \dfrac{\ln(1/W_i)}{\ln r} \leq L. \\[2mm] L, & \text{if } \beta - \dfrac{\ln(1/W_i)}{\ln r} > L. \end{cases} \tag{25}$$

The monotonicity of optimal storage size with respect to importance weight is discussed in the following theorem.

Theorem 2. *Let $(p_1, p_2, \ldots, p_n)$ be a probability distribution and $W = W_1, \ldots, W_n$ be importance weights. L and r are fixed positive integers ($r > 1$). The solution of Problem $\mathcal{P}_2$ meets: $l_i \geq l_j$ if $W_i > W_j$ for $\forall i, j \in \{1, 2, \ldots, n\}$.*

Proof. Refer to the Appendix A. □

This gives rise to a kind of restrictive water-filling, which is presented in Figure 2. Choose a constant β so that $\sum_{i=1}^{n} p_i l_i = T$. The storage size depends on the difference between β and $\frac{\ln(1/W_i)}{\ln r}$. In Figure 2, we obtain that β characterizes the height of water surface, and $\frac{\ln(1/W_i)}{\ln r}$ determines the bottom of the pool. Actually, no storage space is assigned to the data when this difference is less than zero. When the difference is in the interval $[0, L]$, this difference is exactly the storage size. Furthermore, the storage size will be truncated to L bits if the difference is larger than L. Compared with the conventional water-filling, the lowest height of the bottom of the pool is constricted in this restrictive water-filling.

Remark 6. *The restrictive water-filling in Figure 2 is summarized as follows.*

- *For the data with extremely small message importance, $\frac{\ln(1/W_i)}{\ln r}$ is so large that the bottom of the pool is above the water surface. Thus, the storage size of this kind of data is zero.*

- *For the data with small message importance, $\frac{\ln(1/W_i)}{\ln r}$ is large, and therefore the bottom of the pool is high. Thus, the storage size of this kind of data is small.*
- *For the data with large message importance, $\frac{\ln(1/W_i)}{\ln r}$ is small, and therefore the bottom of the pool is low. Thus, the storage size of this kind of data is large.*
- *For the data with extremely large message importance, $\frac{\ln(1/W_i)}{\ln r}$ is so small that the bottom of the pool is constricted in order to truncate the storage size to L.*

Thus, this optimal storage space allocation strategy is a high efficient adaptive storage allocation algorithm for the fact that it can make rational use of all the storage space according to message importance to minimize the RWRE.
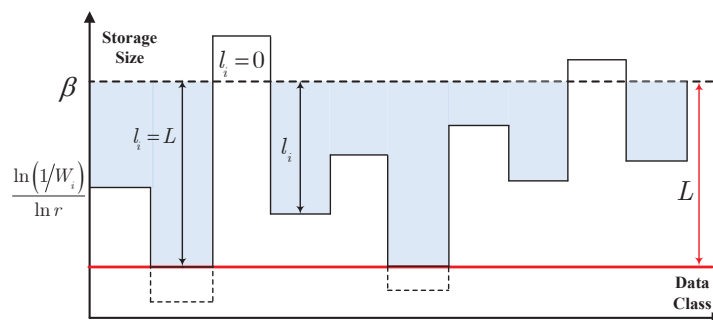

Figure 2. Restrictive water-filling for optimal storage sizes.

This solution can be gotten by means of the recursive algorithm in practice, which is shown in Algorithm 1, where we define an auxiliary function as

$$
f(i, \boldsymbol{W}, \boldsymbol{P}, L, T, r, K_{\min}, K_{\max}) = \begin{cases} L & \text{if } 1 \leq i < K_{\min}. \\ \dfrac{T - \sum_{j=1}^{K_{\min}-1} p_j L}{\sum_{j=K_{\min}}^{K_{\max}} p_j} + \dfrac{\ln W_i}{\ln r} - \dfrac{\sum_{j=K_{\min}}^{K_{\max}} p_j \ln W_j}{\ln r \sum_{j=K_{\min}}^{K_{\max}} p_j} & \text{if } K_{\min} \leq i \leq K_{\max}. \\ 0 & \text{if } K_{\max} < i \leq n. \end{cases} \tag{26}
$$

3.3. Optimal Allocation Strategy in Quantification Storage System

Corollary 2. *For a given maximum available average storage size T ($T \geq 0$), when probability distribution is $(p_1, p_2, \ldots, p_n)$ and the radix is r ($r > 1$), the solution of Problem $\mathcal{P}_3$ is given by*

$$
l_i = \left(\frac{\ln(\ln r) + \ln W_i - \ln \lambda}{\ln r} \right)^+, \tag{27}
$$

where λ is chosen so that $\sum_{i=1}^{n} p_i l_i = T$.

Proof. Let $L \to \infty$ in Corollary 1, the solutions in Equation (19) can be simplified to Equation (27). □

In fact, the optimal storage space allocation strategy in this case can be seen as a kind of water-filling, which gets rid of the constraint on the lowest height of the bottom of the pool.

Algorithm 1 Storage Space Allocation Algorithm

Require:

　　The message importance, $W = \{W_i, i = 1, 2, \ldots, n\}$ (Sort it to satisfy $W_1 \geq W_2 \geq \cdots \geq W_n$)

　　The probability distribution of source, $P = \{p_i, i = 1, 2, \ldots, n\}$

　　The original storage size, L and $L = \{L_i = L, i = 1, 2, \ldots, n\} = \{L, \ldots, L\}$

　　The maximum available average storage size, T

　　The radix, r

　　The auxiliary variables, $K_{\min}, K_{\max}$ (Let $K_{\min} = 1, K_{\max} = n$ as the original values)

Ensure:

　　The compressed storage size, $l = \{l_i, i = 1, \ldots, n\}$

　　Denote this recursive algorithm as $\phi(W, P, L, T, r, K_{\min}, K_{\max})$

 1: 　$l_i' \leftarrow f(i, W, P, L, T, r, K_{\min}, K_{\max})$ for $i = 1, \ldots, n$ 　　　　　$\triangleright$ See Equation (26)

 2: 　**if** $\forall t \in \{1, \cdots, n\}$ such that $0 \leq l_t' \leq L$ and $\sum_{i=1}^{n} p_i l_i' = T$

 3: 　　$l_i \leftarrow l_i'$ for $i = 1, \ldots, n$

 4: **else** 　　**if** $K_{\max} > K_{\min}$

 5: 　　　　$l^{(1)} \leftarrow \phi(W, P, L, T, r, K_{\min}, K_{\max} - 1)$ 　　(Make a recursive call with $K_{\max} \leftarrow K_{\max} - 1$)

 6: 　　　　$\epsilon^{(1)} = D_r(W, L, l^{(1)})$ 　　　　(Calculate the RWRE with $l^{(1)}$) 　　$\triangleright$ See Equation (5)

 7: 　　　　$l^{(2)} \leftarrow \phi(W, P, L, T', r, K_{\min} + 1, K_{\max})$ 　　(Make a recursive call with $K_{\min} \leftarrow K_{\min} + 1$)

 8: 　　　　$\epsilon^{(2)} = D_r(W, L, l^{(2)})$ 　　　　(Calculate the RWRE with $l^{(2)}$) 　　$\triangleright$ See Equation (5)

 9: 　　　　　**if** $\epsilon^{(1)} \leq \epsilon^{(2)}$

10: 　　　　　　$l \leftarrow l^{(1)}$

11: 　　　　　**else**

12: 　　　　　　$l \leftarrow l^{(2)}$

13: 　　　　　**end**

14: 　　　　**else**

15: 　　　　$l_{K_{\min}} \leftarrow (T - \sum_{i=1}^{K_{\min}-1} p_i L)/p_{K_{\min}}$, $l_i \leftarrow L$ when $i < K_{\min}$, $l_i \leftarrow 0$ when $i > K_{\min}$

16: 　　　　**end**

17: **end**

18: **return** l

---

4. Property of Optimal Storage Strategy Based on Message Importance Measure

　　　　Considering that the ideal storage system can capture most of the characteristics of the lossy compression storage model in this paper, we focus on the properties of optimal storage strategy in it in this section for ease of analysis. Specifically, we ignore rounding and adopt l_i in Equation (19) as the optimal storage size of the i-th class in this section. Moreover, we focus on a special kind of the importance weight. Namely, the message importance measure (MIM) is adopted as the importance weight in this part, for the fact that it can effectively measure the cost of the error in data reconstruction in the small-probability event scenarios [23,31].

4.1. Normalized Message Importance Measure

　　　　In order to facilitate comparison under different parameters, the normalized MIM is used and we can write

$$W_i = \frac{e^{\varpi(1-p_i)}}{\sum_{j=1}^{n} e^{\varpi(1-p_j)}}, \tag{28}$$

where ϖ is the importance coefficient, whose selection is discussed in Reference [37]. In fact, the MIM characterizes the user's subjective concern degree of data, and ϖ is an indicator that reflects the user preferences. In practice, the values of ϖ depend on the user preferences. For instance, when ϖ is positive, the user only focuses on the small-probability events, while the large-probability events are focused on when ϖ is negative [33].

Actually, it is easy to check that $0 \leq W_i \leq 1$ for $i = 1, 2, \ldots, n$. Moreover, it is obvious that the sum of those in all event classes is one.

4.1.1. Positive Importance Coefficient

For positive importance coefficient (i.e., $\omega > 0$), let $\alpha_1 = \arg\min_i p_i$ and assume $p_{\alpha_1} < p_i$ for $i \neq \alpha_1$. The derivative of it with respect to the importance coefficient is

$$\frac{\partial W_{\alpha_1}}{\partial \omega} = \frac{\sum_{j=1}^{n} (p_j - p_{\alpha_1}) e^{\omega(2 - p_{\alpha_1} - p_j)}}{\left(\sum_{j=1}^{n} e^{\omega(1 - p_j)} \right)^2} \geq 0. \tag{29}$$

Therefore, W_{α_1} increases as ω increases. In particular, as ω approaches positive infinity, we have

$$\lim_{\omega \to +\infty} W_{\alpha_1} = \lim_{\omega \to +\infty} \frac{e^{\omega(1 - p_{\alpha_1})}}{\sum_{j=1}^{n} e^{\omega(1 - p_j)}} \tag{30}$$

$$= \lim_{\omega \to +\infty} \frac{e^{\omega(1 - p_{\alpha_1})}}{e^{\omega(1 - p_{\alpha_1})} + \sum_{j \neq \alpha_1} e^{\omega(1 - p_j)}} \tag{30a}$$

$$= \lim_{\omega \to +\infty} \frac{1}{1 + \sum_{j \neq \alpha_1} e^{\omega(p_{\alpha_1} - p_j)}} \tag{30b}$$

$$= 1. \tag{30c}$$

Obviously, $\lim_{\omega \to +\infty} W_i = 0$ for $i \neq \alpha_1$.

Remark 7. *As ω approaches positive infinity, the importance weight with the smallest probability is one and others are all zero, which means only a fraction of data almost contains all of the critical information that users care about in the viewpoint of this message importance.*

4.1.2. Negative Importance Coefficient

When the importance coefficient is negative (i.e., $\omega < 0$), let $\alpha_2 = \arg\max_i p_i$ and assume $p_{\alpha_2} > p_i$ for $i \neq \alpha_2$. Its derivative with respect to the importance coefficient is

$$\frac{\partial W_{\alpha_2}}{\partial \omega} = \frac{\sum_{j=1}^{n} (p_j - p_{\alpha_2}) e^{\omega(2 - p_{\alpha_2} - p_j)}}{\left(\sum_{j=1}^{n} e^{\omega(1 - p_j)} \right)^2} \leq 0. \tag{31}$$

Therefore, W_{α_2} decreases as ω increases. In particular, as ω approaches negative infinity, we have

$$\lim_{\omega \to -\infty} W_{\alpha_2} = \lim_{\omega \to -\infty} \frac{e^{\omega(1 - p_{\alpha_2})}}{\sum_{j=1}^{n} e^{\omega(1 - p_j)}} \tag{32}$$

$$= \lim_{\omega \to -\infty} \frac{e^{\omega(1 - p_{\alpha_2})}}{e^{\omega(1 - p_{\alpha_2})} + \sum_{j \neq \alpha_2} e^{\omega(1 - p_j)}} \tag{32a}$$

$$= \lim_{\omega \to -\infty} \frac{1}{1 + \sum_{j \neq \alpha_2} e^{\omega(p_{\alpha_2} - p_j)}} \tag{32b}$$

$$= 1. \tag{32c}$$

Obviously, $\lim_{\omega \to -\infty} W_i = 0$ for $i \neq \alpha_2$.

Remark 8. *As ω approaches negative infinity, the importance weight with the biggest probability is one and others are all zero. If the biggest probability is far from 1, the majority of message importance can also be included in those data with the highest probability, and the corresponding part of the data is not too much.*

4.2. Optimal Storage Size for Each Class

Assume $\tilde{N} = n$ and ignore rounding, due to Equation (23), we obtain

$$l_i = T + \frac{\ln \frac{e^{\omega(1-p_i)}}{\sum_{j=1}^n e^{\omega(1-p_j)}} - \sum_{i=1}^n p_i \ln \frac{e^{\omega(1-p_i)}}{\sum_{j=1}^n e^{\omega(1-p_j)}}}{\ln r} \tag{33}$$

$$= T + \frac{\omega}{\ln r}(\gamma_p - p_i), \tag{33a}$$

where γ_p is an auxiliary variable and it is given by

$$\gamma_p = \sum_{i=1}^n p_i^2. \tag{34}$$

In fact, it is a functional of the minus Rényi entropy of order two, i.e., $\gamma_p = e^{-H_2(P)}$ where $H_2(P)$ is the Rényi entropy $H_\alpha(\cdot)$ when $\alpha = 2$ [38]. Furthermore, we have the following lemma on γ_p.

Lemma 1. *Let $(p_1, p_2, \ldots, p_n)$ be a probability distribution, then we have*

$$\frac{1}{n} \leq \gamma_p \leq 1, \tag{35}$$

$$-\frac{1}{4} \leq \gamma_p - p_i \leq 1. \tag{35a}$$

Proof. Refer to Appendix B. □

Thus, we find $l_i > T$ if $(1/n - p_i)\omega > 0$. Furthermore, we obtain $l_i = T$ when $p_i = \gamma_p$.

Theorem 3. *Let $(p_1, p_2, \ldots, p_n)$ be a probability distribution and $W_i = e^{\omega(1-p_i)}/\sum_{j=1}^n e^{\omega(1-p_j)}$ be the importance weight. The optimal storage sizes in the ideal storage system have the following properties:*

(1) $l_i \geq l_j$ if $p_i < p_j$ for $\forall i, j \in \{1, 2, \ldots, n\}$ when $\omega > 0$;
(2) $l_i \leq l_j$ if $p_i < p_j$ for $\forall i, j \in \{1, 2, \ldots, n\}$ when $\omega < 0$.

Proof. Refer to Appendix C. □

Remark 9. *As noted in [31], the data with smaller probability usually possesses larger importance when $\omega > 0$, while the data with larger probability usually possesses larger importance when $\omega < 0$. Therefore, this optimal allocation strategy makes rational use of all the storage space by providing more storage size for the paramount data and less storage size for the insignificance data. It agrees with the intuitive idea, which is that users generally are more concerned about the data that they need rather than the whole data itself.*

Lemma 2. *Let $(p_1, p_2, \ldots, p_n)$ be a probability distribution and r be radix. L and T are positive integers, and $T < L$. If ω meets $0 \leq T + \omega(\gamma_p - p_i)/\ln r \leq L$, then we have $\tilde{N} = n$.*

Proof. According to Equation (33a) and constraint $0 \leq T + \omega(\gamma_p - p_i)/\ln r \leq L$, we obtain $0 \leq l_i \leq L$ for $\forall i \in \{1, 2, \ldots, n\}$. In this case, $\tilde{N} = n$. □

In fact, when $\varpi \geq 0$, due to Equation (33a) and Lemma 1, we obtain

$$0 \leq T - \frac{\varpi}{4 \ln r} \leq T + \frac{\varpi(\gamma_p - p_i)}{\ln r} \leq T + \frac{\varpi}{\ln r} \leq L. \tag{36}$$

Similarly, when $\varpi < 0$, we have

$$0 \leq T + \frac{\varpi}{\ln r} \leq T + \frac{\varpi(\gamma_p - p_i)}{\ln r} \leq T - \frac{\varpi}{4 \ln r} \leq L. \tag{37}$$

According to Equations (36) and (37), we find $\tilde{N} = n$ always holds if $\max(4 \ln r(T - L), -T/\ln r) \leq \varpi \leq \min(4T \ln r, \ln r(L - T))$.

4.3. Relative Weighted Reconstruction Error

For convenience, we also use $D(\boldsymbol{x}, \varpi)$ to denote the relative weighted reconstruction error (RWRE) $D(\boldsymbol{x}, \boldsymbol{W})$. Due to Equation (7), we have

$$D_r(\boldsymbol{x}, \varpi) = \frac{1}{r^L - 1} \left(\frac{\sum_{i=1}^n p_i e^{\varpi(1-p_i)} r^{L-l_i}}{\sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - 1 \right). \tag{38}$$

If the maximum available average storage size T is zero, then we will have $l_i = 0$ for $i = 1, 2, \ldots, n$. In this case, $D_r(\boldsymbol{x}, \varpi) = 1$. On the contrary, $D_r(\boldsymbol{x}, \varpi) = 0$ when $l_i = L$ for $i = 1, 2, \ldots, n$.

Theorem 4. *$D_r(\boldsymbol{x}, \varpi)$ has the following properties:*

(1) *$D_r(\boldsymbol{x}, \varpi)$ is monotonically decreasing with ϖ in $(0, +\infty)$;*
(2) *$D_r(\boldsymbol{x}, \varpi)$ is monotonically increasing with ϖ in $(-\infty, 0)$;*
(3) *$D_r(\boldsymbol{x}, \varpi) \leq D_r(\boldsymbol{x}, 0) = (r^{L-T} - 1)/(r^L - 1)$.*

Proof. Refer to Appendix D. $\square$

Remark 10. *As shown in Remark 7 and Remark 8, the overwhelming majority of important information will gather in a fraction of data as the importance coefficient increases to negative/positive infinity. Therefore, we can heavily reduce the storage space with extremely small of RWRE with the increasing of the absolute value of the importance coefficient. In fact, this special characteristic of weight reflects the effect of users' preferences. That is, it is beneficial for data compression that the data that users care about is highly clustered. Moreover, when $\varpi = 0$, all the importance weights are the same, which leads to the incompressibility, in a sense, for the fact that there is no special characteristic of weight for users to make rational use of storage space.*

In the following part of this section, we will discuss the cases where $0 \leq T + \varpi(\gamma_p - p_i)/\ln r \leq L$ for $i = 1, \ldots, n$, which means all l_i can be given by Equation (33a) and $n = \tilde{N}$ due to Lemma 2. In this case, substituting Equation (33a) in Equation (7), the RWRE is

$$D_r(\boldsymbol{x}, \varpi) = \frac{e^{\varpi(1-\gamma_p)} r^\Delta}{(r^L - 1) \sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - \frac{1}{r^L - 1}, \tag{39}$$

where $\Delta = L - T$, which characterizes the average compressed storage space of each data.

Since $0 \leq T + \varpi(\gamma_p - p_i)/\ln r \leq L$, we have

$$\begin{cases} \dfrac{\varpi(\gamma_p - p_{\alpha_1})}{\ln r} \leq L - T \leq L + \dfrac{\varpi(\gamma_p - p_{\alpha_2})}{\ln r} & \text{if } \varpi \geq 0. \\[3mm] \dfrac{\varpi(\gamma_p - p_{\alpha_2})}{\ln r} \leq L - T \leq L + \dfrac{\varpi(\gamma_p - p_{\alpha_1})}{\ln r} & \text{if } \varpi < 0. \end{cases} \tag{40}$$

Hence,

$$\delta_1 \le D_r(\boldsymbol{x}, \varpi) \le \delta_2. \tag{41}$$

where

$$\delta_1 = \begin{cases} \dfrac{e^{\varpi(1-p_{\alpha_1})}}{(r^L-1)\sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - \dfrac{1}{r^L-1} & \text{if } \varpi \ge 0, \\[4mm] \dfrac{e^{\varpi(1-p_{\alpha_2})}}{(r^L-1)\sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - \dfrac{1}{r^L-1} & \text{if } \varpi < 0, \end{cases} \tag{42}$$

and

$$\delta_2 = \begin{cases} \dfrac{e^{\varpi(1-p_{\alpha_2})}r^L}{(r^L-1)\sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - \dfrac{1}{r^L-1} & \text{if } \varpi \ge 0. \\[4mm] \dfrac{e^{\varpi(1-p_{\alpha_1})}r^L}{(r^L-1)\sum_{i=1}^n p_i e^{\varpi(1-p_i)}} - \dfrac{1}{r^L-1} & \text{if } \varpi < 0. \end{cases} \tag{43}$$

Theorem 5. *For a given storage system with the probability distribution of data sequence $\boldsymbol{P} = (p_1, p_2, \ldots, p_n)$, let L, r be fixed positive integers ($r > 1$), and ϖ meets $0 \le T + \varpi(\gamma_p - p_i)/\ln r \le L$ for $i = 1, 2, \ldots, n$. For the given least upper bound of the RWRE δ ($\delta_1 \le \delta \le \delta_2$ where δ_1 and δ_1 is defined in Equation (41)), the maximum average compressed storage size of each data $\Delta^*(\delta)$ is given by*

$$\Delta^*(\delta) = \frac{\ln\left(1 + \delta(r^L-1)\right) + L(\varpi, \boldsymbol{P}) - \varpi + \varpi\gamma_p}{\ln r} \tag{44}$$

$$\ge \frac{\ln\left(1 + \delta(r^L-1)\right)}{\ln r}, \tag{44a}$$

where $L(\varpi, \boldsymbol{P}) = \ln\sum_{i=1}^n p_i e^{\varpi(1-p_i)}$, and the equality of Equation (44a) holds if the probability distribution of the data sequence is a uniform distribution or the importance coefficient is zero.

Proof. It is easy to check that $\tilde{N} = n$ according to Lemma 2 for the fact that $0 \le T + \varpi(\gamma_p - p_i)/\ln r \le L$. Let $D(\boldsymbol{x}, \varpi) \le \delta$. By means of Equation (39), we solve this inequality and obtain

$$\Delta \le \frac{\ln\left(1 + \delta(r^L-1)\right) + L(\varpi, \boldsymbol{P}) - \varpi + \varpi\gamma_p}{\ln r} = \Delta^*(\delta), \tag{45}$$

where $L(\varpi, \boldsymbol{P}) = \ln\sum_{i=1}^n p_i e^{\varpi(1-p_i)}$. Then we have the following inequality:

$$\Delta^*(\delta) \overset{(a)}{\ge} \frac{\ln\left(1 + \delta(r^L-1)\right) + \ln e^{\sum_{i=1}^n p_i \varpi(1-p_i)} - \varpi + \varpi\gamma_p}{\ln r} = \frac{\ln\left(1 + \delta(r^L-1)\right)}{\ln r},$$

where (a) follows from Jensen's inequality. Since the exponential function is strictly convex, the equality holds only if $\varpi(1 - p_i)$ is constant everywhere, which means $(p_1, p_2, \ldots, p_n)$ is a uniform distribution or the importance coefficient ϖ is zero. $\square$

Remark 11. *In conventional source coding, the encoding length depends on the entropy of sequence, and a sequence is incompressible if its probability distribution is a uniform distribution [34]. In Theorem 5, the uniform distribution is also the worst case, since the system achieves the minimum compressed storage size. Although the focus is different, they both show that the uniform distribution is detrimental for compression.*

Furthermore, taking $\varpi > 0$ as an example, it is also noted that

$$\Delta^*(\delta) \leq \Delta^*(\delta_2) = L + \frac{\varpi(\gamma_p - p_{\alpha_2})}{\ln r} \leq L, \tag{46}$$

for the fact that $\gamma_p \leq p_{\alpha_2}$. In order to make $\Delta^*(\delta_2)$ approaches L, $\gamma_p - p_{\alpha_2}$ should be as close to zero as possible in the range where $0 \leq T + \varpi(\gamma_p - p_i)/\ln r \leq L$ for $i = 1, 2, \ldots, n$ holds.

When the importance coefficient is constant, for two probability distributions P and Q, if $L(\varpi, P) + \varpi\gamma_p > L(\varpi, Q) + \varpi\gamma_q$, then we will obtain Δ^* in P is larger than that in Q. In fact, $L(\varpi, P)$ is defined as MIM in [30], and $\gamma_p = e^{-H_2(P)}$ [38]. Thus, the maximum average compressed storage size of each data is under the control of MIM and Rényi entropy of order two. For typical small-probability event scenarios where there is an exceedingly small probability, the MIM is usually large, and γ_p is also not small simultaneously with big probability. Therefore, $\Delta^*(\delta)$ is usually large in this case. As a result, much more compressed storage space can be saved in typical small-probability event scenarios while compared to that in uniform probability distribution. Namely, the data can be compressed by means of the characteristic of the typical small-probability events, which may help to improve the design of practical storage systems in big data.

5. Property of Optimal Storage Strategy Based on Non-Parametric Message Importance Measure

In this section, we define the importance weight based on the form of non-parametric message importance measure (NMIM) to characterize the relative weighted reconstruction error (RWRE) [23]. Then, the importance weight of i-th class in this section is given by

$$W_i = \frac{e^{(1-p_i)/p_i}}{\sum_{j=1}^{n} e^{(1-p_j)/p_j}}. \tag{47}$$

Due to Equation (22), the optimal storage size in the ideal storage system by this importance weight is given by

$$
\begin{aligned}
l_i^* &= \min\left(\left\lfloor \frac{T - T_{N_L}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{1}{p_i \ln r} - \frac{1}{\ln r} - \frac{\ln \sum_{j=1}^{n} e^{(1-p_j)/p_j}}{\ln r} - \frac{\sum_{j=1}^{\tilde{N}}(1 - p_{I_j} - p_{I_j} \ln \sum_{j=1}^{n} e^{(1-p_j)/p_j})}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} \right\rfloor^+, L\right) \\
&= \min\left(\left\lfloor \frac{T - T_{N_L}}{\sum_{j=1}^{\tilde{N}} p_{I_j}} + \frac{1}{p_i \ln r} - \frac{\tilde{N}}{\ln r \sum_{j=1}^{\tilde{N}} p_{I_j}} \right\rfloor^+, L\right).
\end{aligned}
\tag{48}
$$

For two probabilities p_i and p_j, if $p_i < p_j$, then we will have $W_i > W_j$. In this case, we obtain $l_i^* \geq l_j^*$ according to Theorem 2.

Assume $\tilde{N} = n$ and ignore rounding, due to Equation (23), we obtain

$$l_i = T + \frac{1}{p_i \ln r} - \frac{n}{\ln r}. \tag{49}$$

Let $0 \leq l_i \leq L$ in this case, we find

$$\frac{1}{n + (L - T)\ln r} \leq p_i \leq \begin{cases} \dfrac{1}{n - T\ln r} & \text{if } n > T\ln r. \\ 1 & \text{if } n \leq T\ln r. \end{cases} \tag{50}$$

Generally, this constraint does not invariably hold, and therefore we usually do not have $\tilde{N} = n$.

For the quantification storage system as shown in $\mathcal{P}_3$ in this section, if the maximum available average storage size satisfies $n \leq T\ln r$, an arbitrary probability distribution will make Equation (50)

hold, which means $\tilde{N} = n$. In this case, substituting Equation (47) in Equation (9), the RWRE can be expressed as

$$D_r(\boldsymbol{x}, \boldsymbol{W}) = \frac{\sum_{i=1}^{n} p_i e^{(1-p_i)/p_i} r^{-l_i}}{\sum_{i=1}^{n} p_i e^{(1-p_i)/p_i}} = e^{n-1-\mathcal{L}(\boldsymbol{P})} r^{-T}, \tag{51}$$

where $\mathcal{L}(\boldsymbol{P}) = \ln \sum_{i=1}^{n} p_i e^{(1-p_i)/p_i}$, which is defined as the NMIM [23].

It is noted $D_r(\boldsymbol{x}, \boldsymbol{W}) = 0$ as T approaches positive infinity. Since $n \leq T \ln r$, we find $D_r(\boldsymbol{x}, \boldsymbol{W}) \leq r^{-1-\mathcal{L}(\boldsymbol{P})}$. Furthermore, since that $\mathcal{L}(\boldsymbol{P}) \geq n - 1$ according to Reference [23], we obtain $D_r(\boldsymbol{x}, \boldsymbol{W}) \leq r^{-n}$. Let $D_r(\boldsymbol{x}, \boldsymbol{W}) \leq \delta$, we have

$$T \geq \frac{n - 1 - \mathcal{L}(\boldsymbol{P}) - \ln \delta}{\ln r}. \tag{52}$$

Obviously, for a given RWRE, the minimum average required storage size for the quantification storage system decreases with increasing of $\mathcal{L}(\boldsymbol{P})$. That is to say, the data with large NMIM will get a large compression ratio. In fact, the NMIM in the typical small-probability event scenarios is generally large according to Reference [23]. Thus, this compression strategy is effective in the typical small-probability event scenarios.

Furthermore, due to Reference [23], $\mathcal{L}(\boldsymbol{P}) \approx \ln p_{\alpha_1} e^{\frac{1-p_{\alpha_1}}{p_{\alpha_1}}}$ when p_{α_1} is small. Hence, for small p_{α_1}, the RWRE in this case can be reduced to

$$D_r(\boldsymbol{x}, \boldsymbol{W}) \approx \frac{e^{n-1/p_{\alpha_1}}}{p_{\alpha_1}} r^{-T}. \tag{53}$$

It is easy to check that $D_r(\boldsymbol{x}, \boldsymbol{W})$ increases as p_{α_1} increases in this case.

6. Numerical Results

We now present numerical results to validate the developed theoretical results in this paper. In this section, we assume all the data is digital, and the exponential distortion measure D_f in Equation (4) is adopted. Furthermore, the relative weighted reconstruction error (RWRE) in Equation (5) is used to characterize the change of total data value before and after the lossy compression based on data value, which represents the total cost of this compression.

6.1. Success Rate of Compressed Storage in General Storage System

This part presents the success rate of compressed storage in the general storage system to show the effectiveness of our method, and it considers the following scenario of data storage.

There are eight categories of data, and the probability distribution of the data category randomly generates in each storing. Moreover, each category of data gets a randomly generated data value, which is in the interval $(0, 100)$. After the lossless conventional data compression, where the data value is assumed to be unchanged, the storage size of each data is a randomly generated number between 10 and 30. The maximum available average storage size T is also varying from 10 to 30 bits. It is considered as a successful data compression when the compressed storage size is not larger than the maximum available storage size. However, when the amount of data to be stored is extremely big, the compressed storage size may still not be enough after the lossless conventional data compression. In this case, the optimal storage space allocation strategy in this paper can be used if a certain amount of data value is allowed to be lost. As a contrast, we also divide up the maximum available storage space equally among all categories of data on the basis of the lossless conventional data compression, which is presented as the equal allocation strategy in Figure 3. Assume that it can also be seen as a successful data compression if the RWRE in this process is less than or equal to the specified amount that can be acceptable by users. For each value of T, this numerical simulation is repeated 10,000

times. The success rate of compressed storage is given by $N_s/10,000$, where N_s is the number of times the successful data compression happens in all the experiments.

Figure 3 shows the relationship between the success rate of compressed storage and the maximum available average storage size T. It is observed that the success rate of conventional data compression is almost one when the available storage size is large ($T > 26$ bits). However, when the available storage size is not big ($T < 26$ bits), the success rate of conventional data compression decreases with decreasing of the maximum available average storage size until it is zero. Furthermore, when a certain amount of data value is allowed to be lost, the success rate can be improved on the basis of the lossless conventional data compression for the same T. More important, the success rate of the optimal allocation strategy is the largest among these three considered compression strategies. For the same maximum available average storage size, the success rates of the optimal allocation strategy and the equal allocation strategy increase as the maximum acceptable RWRE increases. In fact, the success rate of equal allocation strategy is exceedingly close to that of conventional data compression when the maximum acceptable RWRE is small (e.g., 10^{-7}). In general, if a small quantity of total data value is allowed to be lost, our optimal allocation strategy will further improve the performance of data compression on the basis of the lossless conventional data compression.
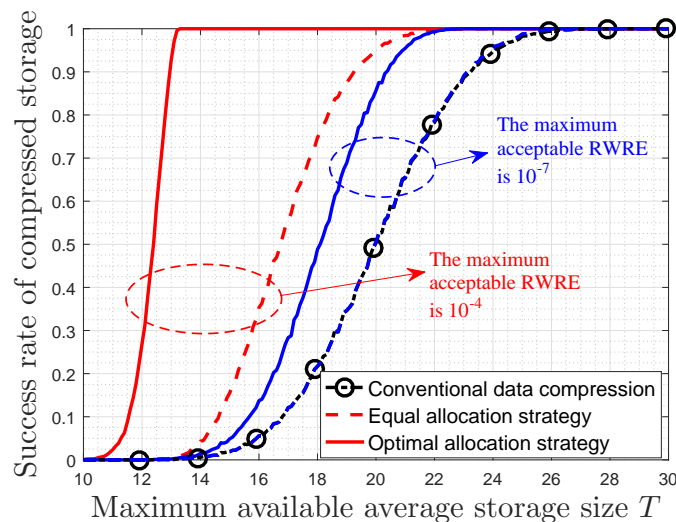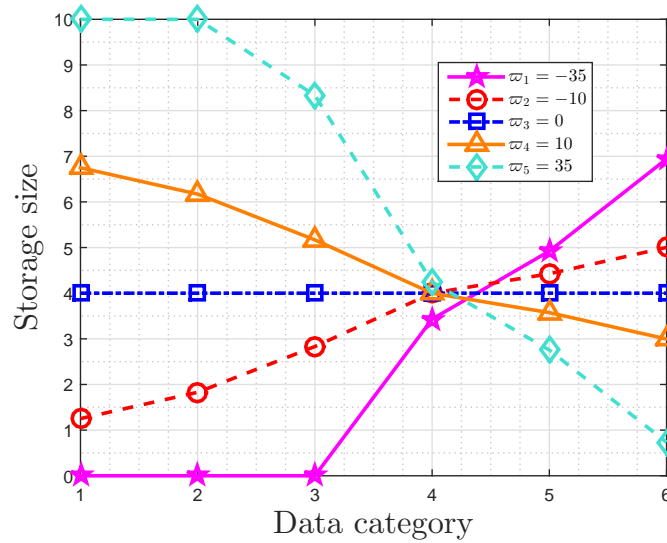

Figure 3. The success rate of compressed storage versus the maximum available average storage size.

6.2. Optimal Storage Size Based on Message Importance Measure in Ideal Storage System

We illustrate the characteristics of optimal storage size based on message importance measure (MIM) in an ideal storage system in this part by means of a broken line graph, which demonstrates the theoretical analyses in Section 4.2. For ease of illustrating, we ignore rounding and the optimal storage size of the i-th class is given by l_i in Equation (19).

The broken line graph of the optimal storage size is shown in Figure 4, when the probability distribution is $P = (0.03, 0.07, 0.1395, 0.2205, 0.25, 0.29)$. In fact, $0.2205 \approx \gamma_P$ and $1/n \approx 0.167$. The maximum available average storage size T is 4 bits, and the original storage size of each data is 10 bits. The importance coefficients are given by $\varpi_1 = -35, \varpi_2 = -10, \varpi_3 = 0, \varpi_4 = 10, \varpi_5 = 35$, respectively. Some observations can be obtained. When $\varpi > 0$, the optimal storage size of the i-th class decreases with the increasing of its probability. On the contrary, the optimal storage size of the i-th class increases as its probability increases when $\varpi < 0$. In addition, the optimal storage size is invariably equal to T ($T = 4$) when $\varpi = 0$. Furthermore, l_i increases as ϖ increases for $i = 1, 2, 3$, and it decreases with ϖ for $i = 5, 6$. For importance coefficients with small absolute values ($\varpi_2, \varpi_3, \varpi_4$), $0 < l_i < L$ holds for $i = 1, 2, \ldots, 6$, and l_4 is extremely close to T ($T = 4$).

Figure 4. Broken line graph of optimal storage size with the probability distribution $(0.03, 0.07, 0.1395, 0.2205, 0.25, 0.29)$, for a given maximum available average storage size $T = 4$ and original storage size $L = 10$.
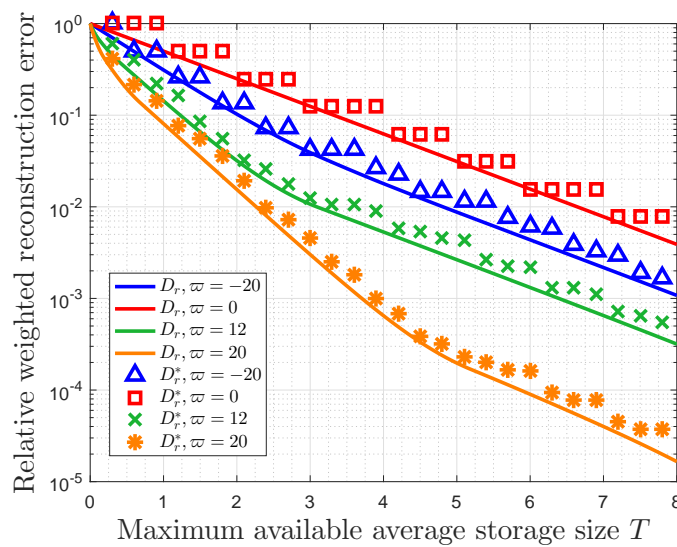
6.3. The Property of the RWRE Based on MIM in Ideal Storage System

Then we focus on the properties of the RWRE. In this part, we will give several numerical results as quintessential examples to validate our theoretical founds in Section 4.3. Without loss of generality, let the original storage size of each data be 16 bits, and the maximum available average storage size T is varying from 0 to 8 bits. Although any range of T can be used, we choose this range to make the results more clear. Moreover, the normalized MIM is adopted to describe the data value that represents the subjective assessment of users.
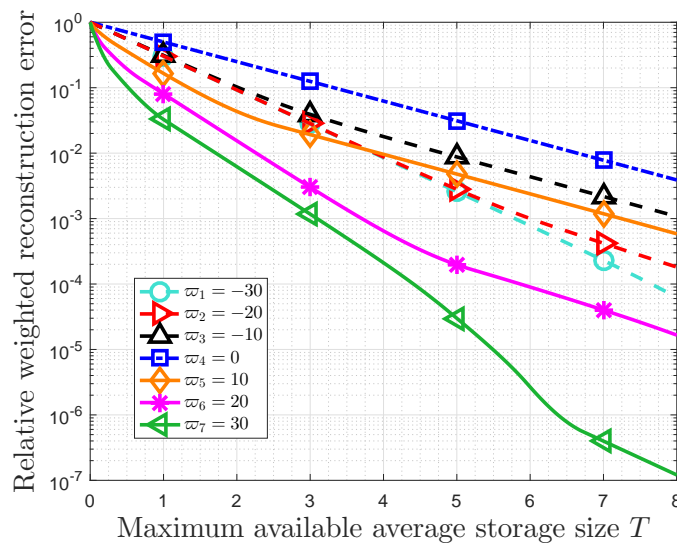
Figures 5 and 6 both present the relationship between the RWRE and the maximum available average storage size with the probability distribution $(0.031, 0.052, 0.127, 0.208, 0.582)$. In fact, the compression ratio is given by T/L in this case, and the RWRE represents the total cost, which measures the compression distortion from the viewpoint of data value. Therefore, these two figures essentially show the trade-off between the compression ratio and the total compressed storage cost.

Figure 5 focuses on the error of RWRE by rounding number with different values of the importance coefficient ϖ ($\varpi = -20, 0, -12, 20$). In Figure 5, the RWRE D_r is acquired by substituting Equation (19) in Equation (38), while the RWRE D_r^* is obtained by substituting Equation (22) in Equation (38). In this figure, D_r^* has a tiered descent as the available average storage size increases, while D_r monotonically decreases with increasing in the available average storage size. Figure 5 also shows that D_r is always less than or equal to D_r^* and they are very close to each other for the same importance coefficient, which means that D_r can be used as the lower bound of D_r^* to reflect the characteristics of D_r^*.

Furthermore, some other observations can be obtained in Figure 6. For the same T, the RWRE increases as ϖ increases when $\varpi < 0$, while the RWRE decreases with increasing of ϖ when $\varpi > 0$. In addition, the RWRE is the largest when $\varpi = 0$. These results prove the validity of Theorem 4. It is also observed that the RWRE always decreases with increasing of T for given ϖ. Furthermore, for any importance coefficient, the RWRE will be 1 if available average storage size is zero. Generally, there is a trade-off between the RWRE and the available storage size, and the results in this paper propose an alternative lossy compression strategy based on message importance.
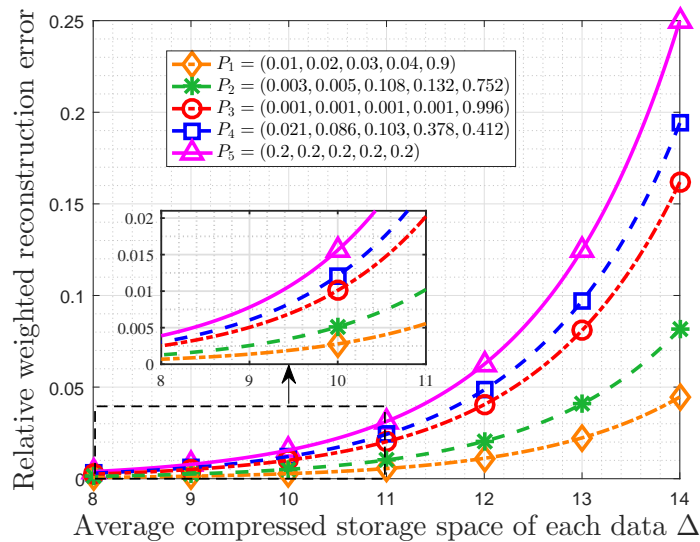
Figure 5. Relative weighted reconstruction error (RWRE) $D_r(\boldsymbol{x}, \varpi)$ versus maximum available average storage size T with the probability distribution $(0.031, 0.052, 0.127, 0.208, 0.582)$ in the case of the value of importance coefficient $\varpi = -20, 0, -12, 20$. D_r is acquired by substituting Equation (19) in Equation (38), while D_r^* is obtained by substituting Equation (22) in Equation (38).

Figure 6. RWRE $D_r(\boldsymbol{x}, \varpi)$ versus maximum available average storage size T with the probability distribution $(0.031, 0.052, 0.127, 0.208, 0.582)$ in the case of the value of importance coefficient $\varpi = -30, -20, -10, 0, 10, 20, 30$.

Then let the importance coefficient ϖ be five and the maximum available average storage size T be varying from two to eight bits. Although any range of T can be used, we choose this range to make the results more clear. In addition, the original storage size is still 16 bits. Furthermore, the average compressed storage space of each data is given by $\Delta = L - T$. In this case, Figure 7 shows that the relationship between the RWRE and the average compressed storage space of each data Δ for different probability distributions. In fact, it can also be seen as reflecting the relationship between the total compressed storage cost and the average saving storage size. The probability distributions and some auxiliary variables are listed in Table 2. In fact, we take these five probability distributions as examples, and $L(\varpi, \boldsymbol{P}) + \varpi e^{-H_2(\boldsymbol{P})}$ of them decreases monotonously. Obviously, all probability distributions satisfy $0 \leq T + \varpi(\gamma_p - p_i)/\ln r \leq L$. It is observed that the RWRE always increases with increasing of Δ for a given probability distribution. Some other observations are also obtained.

For the same Δ, the RWRE of uniform distribution is the largest all the time. Furthermore, if the RWRE is required to be less than a specified value, which is exceedingly common in actual systems in order to make the difference between the raw data and the stored data accepted, the maximum average compressed storage size of each data will increase with increasing of $L(\omega, \boldsymbol{P}) + \omega e^{-H_2(\boldsymbol{P})}$. As an example, when the RWRE is required to be smaller than 0.01, the maximum average compressed storage size of P_1, P_2, P_3, P_4, P_5 of each data is 11.85, 10.97, 9.99, 9.73, 9.36, respectively. In particular, the maximum average compressed storage size of each data in a uniform distribution is the smallest, which suggests the data with a uniform distribution is incompressible from the perspective of optimal storage space allocation based on the data value.

Figure 7. RWRE $D_r(\boldsymbol{x}, \omega)$ vs. average compressed storage size of each data Δ with importance coefficient $\omega = 5$.

Table 2. The auxiliary variables in ideal storage system.

Variable	Probability Distribution	$\omega(\gamma_p - p_{\alpha_1})/\ln r$	$\omega(\gamma_p - p_{\alpha_2})/\ln r$	$L(\omega, P) + \omega e^{-H_2(P)}$
P_1	$(0.01, 0.02, 0.03, 0.04, 0.9)$	5.7924	-0.6276	6.7234
P_2	$(0.003, 0.007, 0.108, 0.132, 0.752)$	4.2679	-1.1350	6.1305
P_3	$(0.001, 0.001, 0.001, 0.001, 0.996)$	7.1487	-0.0287	5.4344
P_4	$(0.021, 0.086, 0.103, 0.378, 0.412)$	2.2367	-0.5838	5.2530
P_5	$(0.2, 0.2, 0.2, 0.2, 0.2)$	0	0	5

6.4. The Property of the RWRE Based on Non-Parametric MIM in a Quantification Storage System

Figure 8 presents the relationship between the RWRE and the maximum available average storage size T for different probability distributions in a quantification storage system, which proves the validity of theoretical results in Section 5. In this part, we use the normalized non-parametric message importance measure (NMIM) to characterize the data value that represents the subjective assessment of users. The probability distributions and some auxiliary variables are listed in Table 3.

Some observations can be obtained. First, the RWRE always decreases with the increasing of the maximum available average storage size for a given probability distribution, and there is a trade-off between the RWRE and the maximum available average storage size. When the maximum available average storage size is small ($T < n/\ln r$), the RWRE decreases largely compared to the case where T is large. In addition, when the maximum available average storage size is large ($T > n/\ln r$), the difference between these RWREs remains the same at the logarithmic Y-axis. In fact, according to Equation (51), this difference between two probabilities in this figure is the difference of NMIM divided by log 10. As an example, the difference between P_1 and P_4 in this figure is 30, which satisfies

this conclusion for the fact that $(\mathcal{L}(P_1) - \mathcal{L}(P_4))/\log 10 \approx 30$. Moreover, the RWRE in P_1 is very close to that in P_2, and the minimum probabilities in these two probability distributions are the same, i.e., $p_{\alpha_1} = 0.007$. It suggests that the data with the same minimum probability will have the same compression performance no matter how the distribution changes, if the minimum probability is very small. In addition, it is also observed that the RWRE decreases as NMIM $\mathcal{L}(P)$ increases for the same T, which means this compression strategy is more effective in the large NMIM cases.
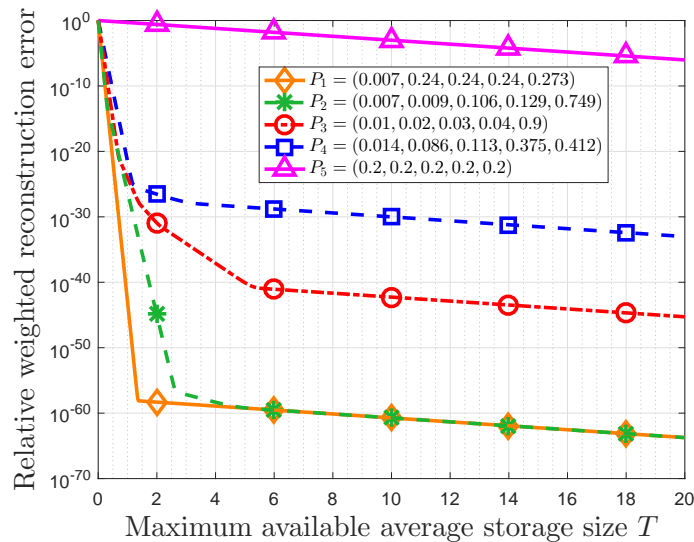

Figure 8. RWRE versus maximum available average storage size T.

Table 3. The auxiliary variables in the quantification storage system.

Variable	Probability Distribution	p_{α_1}	$\mathcal{L}(P)$
P_1	$(0.007, 0.24, 0.24, 0.24, 0.273)$	0.007	136.8953
P_2	$(0.007, 0.009, 0.106, 0.129, 0.749)$	0.007	136.8953
P_3	$(0.01, 0.02, 0.03, 0.04, 0.9)$	0.01	94.3948
P_4	$(0.014, 0.086, 0.113, 0.375, 0.412)$	0.014	66.1599
P_5	$(0.2, 0.2, 0.2, 0.2, 0.2)$	0.2	4.0000

7. Conclusions

In this paper, we focused on the problem of lossy compression storage when the storage size is still not enough after conventional lossless data compression. By means of the message importance to characterize the data value, we define the weighted reconstruction error to describe the total cost in data storage. Based on it, we presented an optimal storage space allocation strategy for digital data from the perspective of data value by the exponential distortion measure, which pursues the least error with respect to the data value for restricted storage size. We gave the solutions by a kind of restrictive water-filling, which presented an alternative way to design an effective storage space allocation strategy. In fact, this optimal allocation strategy prefers to provide more storage size for crucial event classes in order to make the rational use of resources, which agrees with the individuals' cognitive mechanism.

Then, we presented the properties of this strategy based on the message importance measure (MIM) detailedly. It is obtained that there is a trade-off between the relative weighted reconstruction error (RWRE) and available storage size. In fact, if a small quantity of loss of total data value is accepted by users, this strategy will further compress data based on the conventional methods of data compression. Moreover, the compression performance of this storage system improves as the absolute value of importance coefficient increases. This is due to the fact that a fraction of data can contain the overwhelming majority of useful information that exerts a tremendous fascination on users as

the importance coefficient approaches negative/positive infinity, which suggests that the users' interest is highly-concentrated. On the other hand, the probability distribution of event classes also has an effect on the compression results. When the useful information is only highly enriched in a small portion of raw data naturally from the viewpoint of users, such as the small-probability event scenarios, it is obvious that we can compress the data greatly with the aid of these characteristics of distribution. Furthermore, the properties of storage size and RWRE based on non-parametric MIM were also discussed. In fact, the RWRE in the data with a uniform distribution was invariably the largest in any case. Therefore, this paper harbors the idea that the data with uniform information distribution is incompressible from the perspective of optimal storage size allocation based on data value, which is consistent with the well known conclusion in information theory in a sense.

Proposing a more general distortion measure between the raw data and the compressed data, which no longer only applies to digital data, and using it to acquire the high-efficiency lossy data compression systems from the perspective of message importance are of our future interests. In addition, we are also interested in using this optimal storage space allocation strategy in a real application with a real data stream in the future.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of things
MIM	message importance measure
RWRE	relative weighted reconstruction error
NMIM	non-parametric message importance measure

Appendix A. Proof of Theorem 2

In fact, Equation (25) can be rewritten as

$$
l_i = \begin{cases} 0 & \text{if } W_i < e^{-\beta \ln r}. \\ \beta - \dfrac{-\ln W_i}{\ln r} & \text{if } e^{-\beta \ln r} \le W_i \le e^{(L-\beta) \ln r}. \\ L & \text{if } W_i > e^{(L-\beta) \ln r}. \end{cases} \tag{A1}
$$

When $W_i > W_j$, we have

$$
l_i - l_j = \begin{cases} 0 & \text{if } p_i > e^{(L-\beta) \ln r}, \ p_j > e^{(L-\beta) \ln r}. \\ L - \beta - \dfrac{\ln W_j}{\ln r} & \text{if } p_i > e^{(L-\beta) \ln r}, \ e^{-\beta \ln r} \le p_j \le e^{(L-\beta) \ln r}. \\ L & \text{if } p_i > e^{(L-\beta) \ln r}, \ p_j < e^{-\beta \ln r}. \\ \dfrac{\ln W_i - \ln W_j}{\ln r} & \text{if } e^{-\beta \ln r} \le p_i \le e^{(L-\beta) \ln r}, \ e^{-\beta \ln r} \le p_j \le e^{(L-\beta) \ln r}. \\ \beta - \dfrac{-\ln W_i}{\ln r} & \text{if } e^{-\beta \ln r} \le p_i \le e^{(L-\beta) \ln r}, \ p_j < e^{-\beta \ln r}. \\ 0 & \text{if } p_i < e^{-\beta \ln r}, \ p_j < e^{-\beta \ln r}. \end{cases} \tag{A2}
$$

Due to Equation (8b), we obtain that $0 \leq \beta - \frac{-\ln W_i}{\ln r} \leq L$, and therefore $L - \beta - \frac{\ln W_j}{\ln r} \geq 0$. Furthermore, it is easy to check that $\frac{\ln W_i - \ln W_j}{\ln r}$ since that $W_i > W_j$. Thus, $l_i - l_j \geq 0$ if $W_i > W_j$ for $\forall i, j \in \{1, 2, \ldots, n\}$. The proof is completed.

Appendix B. Proof of Lemma 1

(1) For γ_p, it is noted that

$$\sum_{i=1}^{n} p_i^2 = \frac{1}{n} \left(\sum_{i=1}^{n} p_i^2 \sum_{i=1}^{n} 1^2 \right) \geq \frac{1}{n} \left(\sum_{i=1}^{n} p_i \right)^2 = \frac{1}{n}, \tag{A3}$$

where the equality holds only if $(p_1, p_2, \ldots, p_n)$ is a uniform distribution. Moreover,

$$\sum_{i=1}^{n} p_i^2 \leq \sum_{i=1}^{n} p_i = 1, \tag{A4}$$

where the equality holds only if there is only $p_t = 1$ ($t \in \{1, 2, \ldots, n\}$) and $p_k = 0$ for $k \neq t$.

(2) For $\gamma_p - p_i$, we have $\sum_{i=1}^{n} p_i^2 - p_i \leq \sum_{i=1}^{n} p_i^2 \leq 1$. We have equality if and only if $p_t = 1$ and $p_i = 0$ for $i \neq t$. Therefore, we only need to check $\sum_{i=1}^{n} p_i^2 - p_i \geq -1/4$.

First, if $n = 1$, we obtain $\sum_{i=1}^{n} p_i^2 - p_i = 0$.

Second, if $n = 2$, we obtain $\sum_{i=1}^{n} p_i^2 - p_i = 2(p_1 - 3/4)^2 - 1/8$. It is easy to check that $\sum_{i=1}^{n} p_i^2 - p_i \geq -1/8$.

Third, if $n > 3$, we use the method of Lagrange multipliers. Let

$$J(p) = \sum_{j=1}^{n} p_j^2 - p_i - \lambda \left(\sum_{j=1}^{n} p_j - 1 \right). \tag{A5}$$

Setting the derivative to 0, we obtain

$$2p_j^* - \lambda = 0 \quad \text{for } j \neq i \tag{A6}$$

$$2p_j^* - 1 - \lambda = 0 \quad \text{for } j = i. \tag{A6a}$$

Substituting p_j^* in the constraint $\sum_{j=1}^{n} p_j^* = 1$, we have

$$\frac{\lambda(n-1)}{2} + \frac{\lambda + 1}{2} = 1. \tag{A7}$$

Hence, we find $\lambda = 1/n$ and

$$p_j^* = \begin{cases} \dfrac{n+1}{2n} & \text{if } j = i, \\ \dfrac{1}{2n} & \text{if } j \neq i. \end{cases} \tag{A8}$$

In this case, we get

$$\sum_{j=1}^{n} p_j^2 - p_i = \frac{n-1}{4n^2} + \frac{(n+1)^2}{4n^2} - \frac{n+1}{2n} = \frac{-n^2 + n}{4n^2} \geq -\frac{1}{4}. \tag{A9}$$

Thus, Lemma 1 is proved.

Appendix C. Proof of Theorem 3

(1) First, let $p_i < p_j$ when $\varpi > 0$. It is noted that

$$W_i = \frac{e^{\varpi(1-p_i)}}{\sum_{k=1}^n e^{\varpi(1-p_k)}} > \frac{e^{\varpi(1-p_j)}}{\sum_{k=1}^n e^{\varpi(1-p_k)}} = W_j. \tag{A10}$$

Therefore, we find $l_i \geq l_j$ since that $W_i > W_j$, due to Theorem 2.

(2) Second, let $p_i < p_j$ when $\varpi < 0$. It is noted that

$$W_i = \frac{e^{\varpi(1-p_i)}}{\sum_{k=1}^n e^{\varpi(1-p_k)}} < \frac{e^{\varpi(1-p_j)}}{\sum_{k=1}^n e^{\varpi(1-p_k)}} = W_j. \tag{A11}$$

Therefore, we find $l_i \leq l_j$ since that $W_i < W_j$, due to Theorem 2. The proof is completed.

Appendix D. Proof of Theorem 4

We define an auxiliary function as

$$f(\varpi) = \frac{\sum_{i=1} p_i e^{\varpi(1-p_i)} r^{-l_i}}{\sum_{j=1}^n p_j e^{\varpi(1-p_j)}}. \tag{A12}$$

According to Equation (38), it is noted that the the monotonicity of $D_r(x, \varpi)$ with respect to ϖ is the same with that of $f(\varpi)$.

Without loss of generality, let l_i of p_i be

$$l_i = \begin{cases} L & \text{if } i = 1, 2, \ldots, t_1, \\ \dfrac{\ln(\ln r) + \ln W_i - \ln \lambda}{\ln r} & \text{if } i = t_1 + 1, \ldots, t_2, \\ 0 & \text{if } i = t_2 + 1, t_2 + 2, \ldots, n, \end{cases} \tag{A13}$$

where λ is given by Equation (20) where $\{T_j, j = 1, \ldots, \tilde{N}_L\} = \{1, 2, \ldots, t_1\}$ and $\{I_j, j = 1, \ldots, \tilde{N}\} = \{t_1 + 1, \ldots, t_2\}$.

The derivative of l_i with respect to ϖ is given by

$$l_i' = \begin{cases} \dfrac{\sum_{k=t_1+1}^{t_2} p_k(p_k - p_i)}{\ln r(\sum_{k=t_1+1}^{t_2} p_k)} & \text{if } i = t_1 + 1, \ldots, t_2. \\ 0 & \text{else.} \end{cases} \tag{A14}$$

Hence,

$$f'(\varpi) = \frac{\sum_i \sum_j p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i}(p_j - p_i - l_i' \ln r)}{\left(\sum_j p_j e^{\varpi(1-p_j)}\right)^2} = \frac{F_1 + F_2}{\left(\sum_j p_j e^{\varpi(1-p_j)}\right)^2}, \tag{A15}$$

where $F_1 = \sum_i \sum_j p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i}(p_j - p_i)$ and $F_2 = \sum_i \sum_j p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i}(-l_i' \ln r)$.

(1) When $\varpi > 0$, we have

$$F_1 = \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i} (p_j - p_i) + \sum_{p_j > p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i} (p_j - p_i) \tag{A16}$$

$$\leq \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i) + \sum_{p_j > p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i} (p_j - p_i) \tag{A16a}$$

$$= \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i) + \sum_{p_i > p_j} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_i - p_j) \tag{A16b}$$

$$= \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i + p_i - p_j) \tag{A16c}$$

$$= 0. \tag{A16d}$$

In fact, if $p_i > p_j$, then we will have $l_i \leq l_j$ due to Theorem 3. Thus, $r^{-l_i}(p_j - p_i) \leq r^{-l_j}(p_j - p_i)$ in this case. With taking $p_i p_j e^{\varpi(2-p_i-p_j)} \geq 0$ into account, we have Equation (A16a). Furthermore, Equation (A16b) is obtained by exchanging the notation of subscript in the second item.

For $t_1 < i \leq t_2$ and $1 \leq j \leq n$, we have

$$F_2 = \sum_{i=t_1+1}^{t_2} \sum_{j=1}^{n} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i} (-l_i' \ln r) \tag{A17}$$

$$= \sum_{i=t_1+1}^{t_2} \sum_{j=1}^{n} p_i p_j e^{\varpi(1-p_j)-\ln \ln r + \ln \lambda} (-l_i' \ln r) \tag{A17a}$$

$$= \sum_{j=1}^{n} \left(p_j B_j \left(\sum_{i=t_1+1}^{t_2} p_i (-l_i' \ln r) \right) \right) \tag{A17b}$$

$$= \sum_{j=1}^{n} \left(p_j B_j \left(\frac{\sum\limits_{i=t_1+1}^{t_2} p_i^2 \sum\limits_{k=t_1+1}^{t_2} p_k - \sum\limits_{k=t_1+1}^{t_2} p_k^2 \sum\limits_{i=t_1+1}^{t_2} p_i}{\sum\limits_{k=t_1+1}^{t_2} p_k} \right) \right) \tag{A17c}$$

$$= 0, \tag{A17d}$$

where $B_j = \exp\{\varpi(1 - p_j) - \ln \ln r + \ln \lambda\}$.

Based on the discussions above, we have

$$f'(\varpi) = \frac{F_1 + F_2}{\left(\sum_{i=1}^{n} p_i e^{\varpi(1-p_i)} \right)^2} \leq 0. \tag{A18}$$

Since that $f'(\varpi) \leq 0$ when $\varpi > 0$, $D_r(x, \varpi)$ is monotonically decreasing with ϖ in $(0, +\infty)$.

(2) Similarly, when $\varpi < 0$, if $0 < p_j < p_i$, then we will have $l_i > l_j$ due to Theorem 3. Thus, $r^{-l_i}(p_j - p_i) \geq r^{-l_j}(p_j - p_i)$ in this case. With taking $p_i p_j e^{\varpi(2-p_i-p_j)} \geq 0$ into account, we have

$$F_1 \geq \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i) + \sum_{p_j > p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_i} (p_j - p_i) \tag{A19}$$

$$= \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i) + \sum_{p_i > p_j} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_i - p_j) \tag{A19a}$$

$$= \sum_{p_j < p_i} p_i p_j e^{\varpi(2-p_i-p_j)} r^{-l_j} (p_j - p_i + p_i - p_j) \tag{A19b}$$

$$= 0, \tag{A19c}$$

where Equation (A19a) is obtained by exchanging the notation of subscript in the second item.

In addition, F_2 is still given by Equation (A17), and $F_2 = 0$. As a result, $f'(\varpi) \geq 0$ when $\varpi < 0$. Therefore $D_r(\boldsymbol{x}, \varpi)$ is monotonically increasing with ϖ in $(-\infty, 0)$.

(3) When $\varpi = 0$, the storage size l_i for $i = 1, 2, \ldots, n$ will be all equal to T, and therefore $D_r(\boldsymbol{x}, 0) = (r^{L-T} - 1)/(r^L - 1)$. Based on the discussion in (1) and (2), we obtain $D_r(\boldsymbol{x}, \varpi) \leq D_r(\boldsymbol{x}, 0)$. The proof is completed.

References

1. Chen, M.; Mao, S.; Zhang, Y.; Leungm, V.C. Definition and features of big data. In *Big Data: Related Technologies, Challenges and Future Prospects*; Springer: New York, NY, USA, 2014; pp. 2–5.
2. Cai, H.; Xu, B.; Jiang, L.; Vasilakos, A. IoT-based big data storage systems in cloud computing: Perspectives and challenges. *IEEE Internet Things J.* **2017**, *4*, 75–87. [CrossRef]
3. Hu, H.; Wen, Y.; Chua, T.; Li, X. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access* **2014**, *2*, 652–687.
4. Dong, D.; Herbert, J. Content-aware partial compression for textual big data analysis in hadoop. *IEEE Trans. Big Data* **2018**, *4*, 459–472. [CrossRef]
5. Park, J.; Park, H.; Choi, Y. Data compression and prediction using machine learning for industrial IoT. In Proceedings of the IEEE International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 10–12 January 2018; pp. 818–820.
6. Geng, D.; Zhang, C.; Xia, C.; Xia, X.; Liu, Q.; Fu, X. Big data-based improved data acquisition and storage system for designing industrial data platform. *IEEE Access* **2019**, *7*, 44574–44582. [CrossRef]
7. Nalbantoglu, Ö.; Russell, D.; Sayood, K. Data compression concepts and algorithms and their applications to bioinformatics. *Entropy* **2010**, *12*, 34–52. [CrossRef]
8. Cao, X.; Liu, L.; Cheng, Y.; Shen, X. Towards energy-efficient wireless networking in the big data era: A survey. *IEEE Commun. Surv. Tutor.* **2017**, *20*, 303–332. [CrossRef]
9. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
10. Oohama, Y. Exponential strong converse for source coding with side information at the decoder. *Entropy* **2018**, *20*, 352. [CrossRef]
11. Pourkamali-Anaraki, F.; Becker, S. Preconditioned data sparsification for big data with applications to pca and k-means. *IEEE Trans. Inf. Theory* **2017**, *63*, 2954–2974. [CrossRef]
12. Aguerri, I.E.; Zaidi, A. Lossy compression for compute-and-forward in limited backhaul uplink multicell processing. *IEEE Trans. Commun.* **2016**, *64*, 5227–5238. [CrossRef]
13. Cui, T.; Chen, L.; Ho, T. Distributed distortion optimization for correlated sources with network coding. *IEEE Trans. Commun.* **2012**, *60*, 1336–1344. [CrossRef]
14. Ukil, A.; Bandyopadhyay, S.; Sinha, A.; Pal, A. Adaptive Sensor Data Compression in IoT systems: Sensor data analytics based approach. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 5515–5519.
15. Zhong, J.; Yates, R.D.; Soljanin, E. Backlog-adaptive compression: Age of information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 566–570.
16. Elkan, C. The foundations of cost-sensitive learning. In Proceedings of the the Seventeenth International Joint Conference on Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001; pp. 973–978.
17. Zhou, Z.; Liu, X. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 63–77. [CrossRef]
18. Lomax, S.; Vadera, S. A survey of cost-sensitive decision tree induction algorithms. *ACM Comput. Surv.* **2013**, *45*, 16:1–16:35. [CrossRef]
19. Masnick, B.; Wolf, J. On linear unequal error protection codes. *IEEE Trans. Inf. Theory* **1967**, *3*, 600–607. [CrossRef]
20. Sun, K.; Wu, D. Unequal error protection for video streaming using delay-aware fountain codes. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017; pp. 1–6.
21. Feldman, D.; Schmidt, M.; Sohler, C. Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering. In Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 6–8 January 2013; pp. 1434–1453.

22. Tegmark, M.; Wu, T. Pareto-optimal data compression for binary classification tasks. *Entropy* **2020**, *22*, 7. [CrossRef]

23. Liu, S.; She, R.; Fan, P.; Letaief, K.B. Non-parametric message important measure: Storage code design and transmission planning for big data. *IEEE Trans. Commun.* **2018**, *66*, 5181–5196. [CrossRef]

24. Ivanchev, J.; Aydt, H.; Knoll, A. Information maximizing optimal sensor placement robust against variations of traffic demand based on importance of nodes. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 714–725. [CrossRef]

25. Kawanaka, T.; Rokugawa, S.; Yamashita, H. Information security in communication network of memory channel considering information importance. In Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; pp. 1169–1173.

26. Li, M.; Zuo, W.; Gu, S.; Zhao, D.; Zhang, D. Learning convolutional networks for content-weighted image compression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3214–3223.

27. Zhang, X.; Hao, X. Research on intrusion detection based on improved combination of K-means and multi-level SVM. In Proceedings of the IEEE International Conference on Communication Technology (ICCT), Chengdu, China, 27–30 October 2017; pp. 2042–2045.

28. Li, M. Application of cart decision tree combined with pca algorithm in intrusion detection. In Proceedings of the IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 38–41.

29. Beasley, M.S.; Carcello, J.V.; Hermanson, D.R.; Lapides, P.D. Fraudulent financial reporting: Consideration of industry traits and corporate governance mechanisms. *Account. Horiz.* **2000**, *14*, 441–454. [CrossRef]

30. Fan, P.; Dong, Y.; Lu, J.; Liu, S. Message importance measure and its application to minority subset detection in big data. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016; pp. 1–6.

31. She, R.; Liu, S.; Wan, S.; Xiong, K.; Fan, P. Importance of small probability events in big data: Information measures, applications, and challenges. *IEEE Access* **2019**, *7*, 100363–100382. [CrossRef]

32. She, R.; Liu, S.; Fan, P. Recognizing Information Feature Variation: Message Importance Transfer Measure and Its Applications in Big Data. *Entropy* **2018**, *20*, 401. [CrossRef]

33. Liu, S.; Dong, Y.; Fan, P.; She, R.; Wan, S. Matching users' preference under target revenue constraints in data recommendation systems. *Entropy* **2019**, *21*, 205. [CrossRef]

34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.

35. Aggarwal, C.C. *Data Classification: Algorithms and Applications*; CRC Press: Boca Raton, FL, USA, 2014.

36. Salvador–Meneses, J.; Ruiz–Chavez, Z.; Garcia–Rodriguez, J. Compressed kNN: k-nearest neighbors with data compression. *Entropy* **2019**, *21*, 234. [CrossRef]

37. She, R.; Liu, S.; Dong, Y.; Fan, P. Focusing on a probability element: Parameter selection of message importance measure in big data. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 20–26 May 2017; pp. 1–6.

38. Van Erven, T.; Harremoës, P. Rényi divergence and kullback-leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]