Journal of Multimedia and Information System VOL. 3, NO. 2, June 2016(pp. 21-26): ISSN 2383-7632(Online)
http://dx.doi.org/10.9717/JMIS.2016.3.2.21

Noisy Data Aggregation with Independent Sensors: Insights and Open Problems

Tatsuto Murayama[1,*], Peter Davis[2]

Abstract

Our networked world has been growing exponentially fast. The explosion in volume of machine-to-machine (M2M) transactions threatens to exceed the transport capacity of the networks that link them. Therefore, it is quite essential to reconsider the tradeoff between using many data sets versus using good data sets. We focus on this tradeoff in the context of the quality of information aggregated from many sensors in a noisy environment. We start with a basic theoretical model considered in the famous ``CEO problem'' in the field of information theory. From a point of view of large deviations, we successfully find a simple statement for the optimal strategies under the limited network capacity condition. Moreover, we propose an open problem for a sensor network scenario and report a numerical result.

Key Words: Noisy Sensing, Data Aggregation, Network Capacity, Large Deviations.

I. INTRODUCTION

How should we aggregate data from the many sensors around the world? What is the optimal way, if any, to do that? Or more precisely, what exactly are the suitable measures for such new optimization problems? These are the issues that we focus on in this paper [1]. The background of our interest comes from the recent growth of the new digital world in which we can get data from huge numbers of sensors distributed worldwide. Machine-to-machine (M2M) communications are said to consume as much as 70 percent of information flows in the Internet—or perhaps we should say, the Internet of Things (IoT). This expanding digital universe offers so much data and information about our environment that we are able to understand what is happening around us even though we are not always staring at it right in front of us. Instead, loyal agents, sensors equipped with digital eyes and ears engineered by us, can keep gathering relevant information on our behalf [2]. However, this situation causes a new kind of problem. That is, since sensors have become cheap to buy or create, we might not have sufficient network capacity to carry the data streams generated by these diligent devices. Therefore, it is crucial to investigate how to manage a large number of sensors under realistic constraints on network resources, in particular, network bandwidth

To this aim, in this paper, we present a theoretical approach which is based on the Gaussian approximation. This is the most intuitive approach for analyzing and understanding the fundamental tradeoff, that is, using many sensors involves the difficult problem of collecting and merging data, whereas using very few sensors only offers little about the subject. The Gaussian analysis captures this sort of tradeoff quantitatively well with a rather simple and easy calculation. Moreover, we also show some non-trivial behavior of such data aggregation tasks, from a point of view of a new model for ad-hoc sensor networks.

Manuscript received March 06, 2016; Revised July 20, 2016; Accepted July 28, 2016. (ID No. JMIS-2016-0003)
Corresponding Author (*): Tatsuto Murayama, Gofuku 3190, Toyama-shi, 930-8555, Japan, +81-76-445-6746, murayama@eng.u-toyama.ac.jp.
[1]Graduate School of Science and Engineering, University of Toyama, Toyama, Japan, murayama@eng.u-toyama.ac.jp
[2]Telecognix Corporation, Kyoto, Japan, davis@telecognix.com.

II. SYSTEM MODEL

In this paper, we consider a binary model for the sensing system. We assume that the state of the sensing target $X(t)$ and the corresponding sensing result $Y_a(t)$ are all binary symbols for sensing event label $t = 1, 2, \cdots, T$ and sensor label $a = 1, 2, \cdots, L$. We let $x(t)$ be a realization of the random variable $X(t)$ and $y_a(t)$ be a corresponding realization of the random variable $Y_a(t)$. In this paper, we assume that $X(t)$ is a collection of the Bernoulli (1/2) random variables. That is, the probability of getting $x(t) = 1$ is always 1/2 for any t, and similarly for $x(t) = 0$. This simplest setup implies that there is no redundancy in the information.

2.1. Noisy Sensing

Since we assume a certain level of noise, the individual values $y_a(t)$ could be different for different values of the sensor label a. The simplest model for the noisy sensing would be a stochastic process defined to be

$$P(y_a(t)|\ x(t)) = \begin{cases} 1-p & \text{if } x(t) = y_a(t) \\ p & \text{otherwise} \end{cases},$$

which is often called the binary symmetric channel (BSC) in the field of information theory. Notice that we assumed that the flip probability is the same for all sensors.

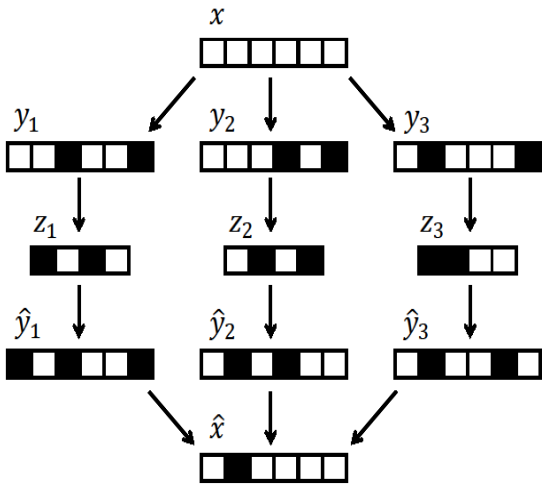

Fig.1. The communications system model for noisy sensing with separate encoding/decoding ansatz, where the system level aggregation is done by the majority vote.

2.2. Separate Encoding

Now, the sensors encode the noisy data bits $y_a(t)$ of length T into their codewords, say $z_a(s)$, of length S that are also binary sequences. In such a case, the individual data rate for the encoding is $R = S/T$. We assume that the individual data rate does not depend on the value of the label a. Therefore, the combined data rate is simply given by the formula $C = RL$. In the CEO problem, a famous model for sensing and communications tasks, the encoding is done independently at every sensor a, while we have no such restrictions on the decoders. This separate encoding assumption is quite natural for a collection of independent sensors, since mutual communications tasks require some computational resources.

2.3. Separate Decoding

As for the decoding process, we use notations $\hat{y}_a(t)$ for the reconstructions of noisy data bits $y_a(t)$. Contrary to the spirit of the CEO problem, we do not focus further on the optimal joint decoding with $\hat{y}_a(t)$. Instead, we restrict ourselves on considering the practical scenario in which the estimate for the original data bits, say $\hat{x}$, is determined by the collection of $\hat{y}_a(t)$ that are independently decoded by the corresponding decoders.

2.4. Data Aggregation Tasks

Lastly, the central computer then uses the reconstructed data $\hat{y}_a(t)$ to calculate the estimate $\hat{x}$ for the original bit x. Here we assume that the reproductions $\hat{y}_a(t)$ have the same distortion level, corresponding to the assumption that all sensor agents have the same ability for encoding/decoding tasks. This is called the exchangeable sensor ansatz.

Hereafter, what we call the distortion will be the Hamming distortion, defined to be

$$d(x(t), y_a(t)) = \begin{cases} 0 & \text{if } x(t) = y_a(t) \\ 1 & \text{otherwise} \end{cases}$$

The distortion measure is so far defined on a bit-by-bit basis. However, it is an easy matter to extend the above definition to the whole sequence of bits. The distortion between the sequences would be the average value of such distortion bits $\langle d(X(t), Y_a(t)) \rangle$. Since we impose the exchangeable sensor ansatz, as is mentioned before, it is known that the optimal estimate for the original bit is nothing but the majority vote,

$$\hat{x}(t) = \begin{cases} 1 & \text{if } \langle \hat{y}_a(t) \rangle \geq 0.5 \\ 0 & \text{otheriwse} \end{cases}$$

Together with the bit-wise calculation for each $\hat{y}_a(t)$, which can be done sequentially, we can easily give the overall estimate for the original data bits $x(t)$.

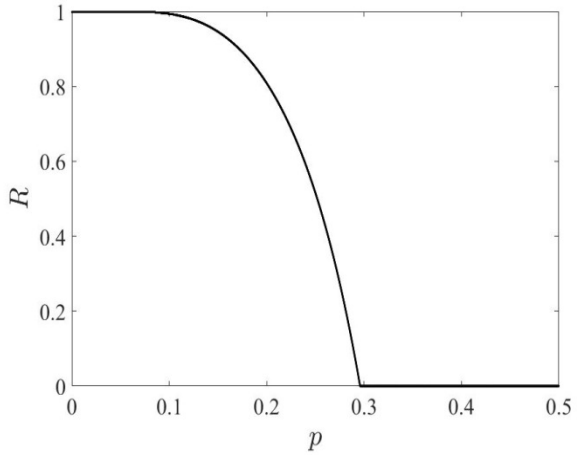

Fig.2. The optimal data rates for lossy coding given a certain noise level. The combined data rate is fixed to a large number.

III. OPTIMALITY MEASURE

Next, we introduce and define the system-level performance based on what we call the expected bit error rate, or simply BER, in information theory. This measure could be written by $P(X(t) \neq \hat{X}(t))$ or more explicitly

$$I_p(R) = \lim_{C \to \infty} \frac{-1}{C} \log_e P\left(X(t) \neq \hat{X}(t)\right) \quad (1)$$

for a given pair of noise level p and individual data rate R [3]. This indicates that overall systematic errors reduce exponentially fast when the combined data rate, not an individual one, tends to infinity. Since our exchangeable sensor ansatz yields the Bernoulli process in the analysis, it is an easy matter to check that for $0 < R \leq 1$,

$$I_p(\mathrm{R}) = \frac{\alpha(p,R)^2}{2R\big(1-\alpha(p,R)\big)\big(1+\alpha(p,R)\big)} \quad (2)$$

with

$$\alpha(p,R) = (1-2p)\big(1-2D(R)\big) \quad (3)$$

where $D(R)$ denotes the performance of our encoder and decoder pair [4]. The preceding formula for the exponential rate of decay is based on the simple Gaussian approximation, which enables us to qualitatively capture the system level behavior of a collective system of this kind. Below, Figure 3 shows the numerical evaluation for the tradeoff between the binary strategic options with the data rates $R \to 0$ and $R = 1$, respectively. In the next section, we see what happens if the scheme is applied together with practical heuristics.

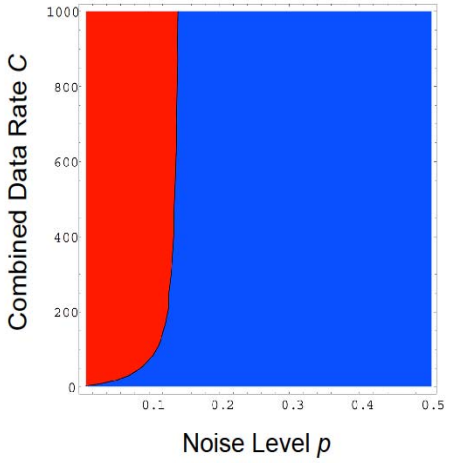

Fig.3. Schematic representation of threshold behavior of optimal data rates for individual sensors. Red colored region corresponds to optimal data rate $R^* = 1$, while blue colored region indicates $R^* \to 0$ if we only have the binary options.

IV. EXPERIMENTS

We first consider a class of practical encoders, which are based on low-density parity check error-correcting codes and message passing technique for lossy compression [5,6]. The simplest heuristic algorithm here would be one that is based on the so-called Thouless-Anderson-Palmer's approach in physics, or what we call reinforcement belief propagation in terms of information theory. In our notations, the procedure can be written as below. Write a set of newly defined variables as $m_{st}^a(j)$, $\hat{m}_{st}^a(j)$ for $j = 1, 2, \cdots$. Then, we find

$$m_{st}^a(j) = \tanh\left(\sum_{t' \in M(s)\setminus t} \tanh^{-1} \hat{m}_{st'}^a(j) + \tanh^{-1} \gamma m_s(j)\right)$$

$$\hat{m}_{st}^a(j+1) = \tanh\big(\beta J(t)\big) \prod_{s' \in L(t)\setminus s} m_{s't}(j)$$

with a posterior approximation

$$m_s^a(j) = \tanh\left(\sum_{t \in M(s)} \tanh^{-1} \hat{m}_{st}^a(j) + \tanh^{-1} \gamma m_s(j)\right)$$

where $J(t)$ represents the antipodal translation of $x(t)$. These equations give an iterative procedure to get a collection of $z_a(s)$ that could be calculated from the Boolean translation of $m_s^a(j)$ for the steps j large enough [7].

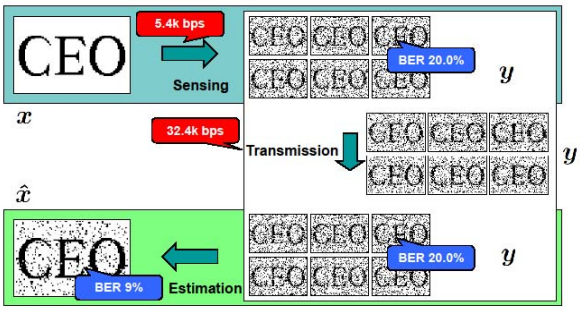

Fig.4. Example of noisy data aggregation without coding the individual sequences. The data rate is $R = 1$.

To see how this works, we consider some simple examples which may demonstrate the potential benefit of our strategy. Figure 4 denotes the noisy data aggregation without any coding techniques. In other words, we set the value of individual data rate to be one. In this setup, the bit error rate for the final estimate would be 9%.

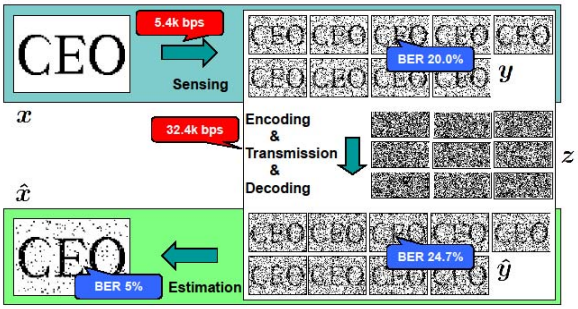

Fig.5. Example of noisy data aggregation with coding the individual sequences. The data rate is R=2/3.

On the other hand, Figure 5 represents the benefit from our sensing data aggregation strategy with some coding techniques. Here we find that the final bit error rate is as small as 5%. These results indicate that there exists a certain noise level in which our large scale data aggregation scheme does outperform the standard type of information gathering without the method of lossy data compression.

V. OPEN PROBLEMS

In this paper, we considered insights from a point of view of large deviations, where one observes a non-trivial tradeoff between the collection of many data and the collection of good data. Here we defined the system level optimality measure (1) as the exponential rate of decay of expected errors for the estimation, assuming that the system size would be large enough. According to the traditional and pedagogical Gaussian approach, one observes a kind of ``phase transition'' of optimal data rate with respect to the noise level. In other words, there is a critical point for the noise level, beyond which it is better for us to deploy as many sensors as possible. This is called the second order transition, implying the continuity of the optimal data rates for various noise levels.

Recently, we have shown that there may be another kind of transition if one considers sensor networks for which the combined data rate is not fixed but actually increases when we deploy more sensors. In this case, the behavior of the optimal data rate drastically changes.

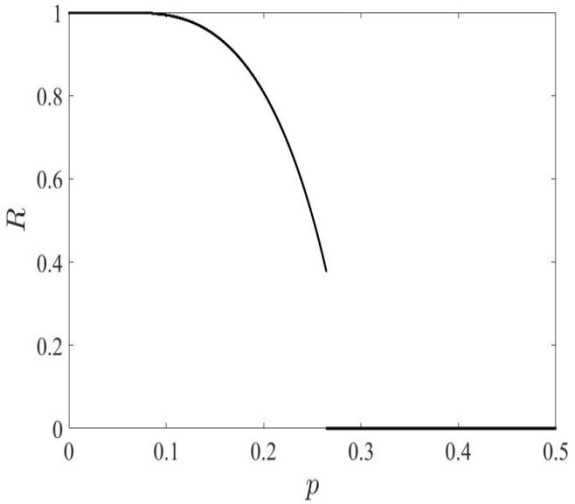

Fig.6. The optimal data rates for lossy coding given a certain noi se level in a sensor network where the combined data rate increas es at a certain rate if one deploys more sensors. Here we used $\gamma = 1.0 \times 10^{-3}$.

In this case, we consider the dynamical capacity constraint $CL^{\gamma} = RL$, instead of $C = RL$, with a parameter γ that controls the rate of increase of the total flow per target sink. While the condition $\gamma = 0$ retrieves the conventional CEO model, we could analyze more generic situations $0 \leq \gamma \leq 1$ by considering the same (1) as the optimality measure for the system [8]. As you can see in Figure 6 the

optimal data rate is not a continuous function with respect to the noise level any more. Instead, it is discontinuous at the point around the critical point we have discussed in this paper. This type of transition is often called a first order transition in the field of statistical mechanics. At this point, however, this is nothing but a naïve conjecture based on the Gaussian approximation (2) with (3). We are now carrying out a large scale simulation to verify this phenomenon.

In this paper, we consider a binary model for the sensing system. We assume that the state of the sensi ng target $X(t)$ and the corresponding sensing result $Y_a(t)$ are all binary symbols for sensing event label $t = 1, 2, \cdots, T$ and sensor label $a = 1, 2, \cdots, L$. We let $x(t)$ be a realization of the random variable $X(t)$ and $y_a(t)$ be a corresponding realization of the random v ariable $Y_a(t)$. In this paper, we assume that $X(t)$ is a collection of the Bernoulli(1/2) random variables. That is, the probability of getting $x(t) = 1$ is always 1/2 f or any t, and similarly for $x(t) = 0$. This simplest se tup implies that there is no redundancy in the inform ation.

Acknowledgement

We would like to thank Masato Tajima and Koji Okino for useful discussions. This work was in part supported by JSPS KAKENHI Grant Numbers 24650073, 26120516.

REFERENCES

[1] D. Culler and H. Mulder, "Smart sensors to network the world," *Scientific American*, vol. 290, no. 6, pp. 84-91, 2004.

[2] J. Gantz and D. Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," available at *http://www.emc.com*, 2012.

[3] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO Problem," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887-902, 1996.

[4] T. Murayama and P. Davis, "Universal Behavior in large-scale aggregation of independent noisy observations," *EPL*, vol. 87, no. 4, 48003, 2009.

[5] R. G. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21-28, 1962.

[6] E. Martinian and J. S. Yedidia, "Iterative Quantization Using Codes On Graphs," in *Proceedings of the 41st Annual Allerton Conference on Communications, Control, and Computing*, Urbana, IL, Oct. 2003.

[7] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Physical Review E*, vol. 69, 035105(R), 2004.

[8] T. Murayama, K. Okino, M. Tajima, and P. Davis, "Aggregation Principle for Independent Noisy Observations: A Scaling-law Perspective," presented at *the 2nd Korea-Japan Joint Workshop on Complex Communication Sciences*, Okinawa, Japan, Oct. 2013.

Authors

Tatsuto Murayama has been a Lecturer of Graduate School of Science and Engineering at University of Toyama since 2013. He received the B.S., M.S., and Ph.D. degrees from Tokyo Institute of Technology. From 2002 to 2004, he was a special postdoctoral researcher at RIKEN Brain Science Institute. From 2004 to 2013, he was a researcher at NTT Communication Science Laboratories. His research interests include data compression, error correction, and collective behaviors in the area of computer science.

Peter Davis was educated at the University of Queensland, Brisbane, in degree programs of B.Sc.(Hons) and Ph.D. In 1987, he joined Advanced Telecommunication Research Institute International (ATR), Kyoto, where he was with the ATR Optical and Radio Communications Research Laboratories from 1987 to 1996, and with the ATR Adaptive Communications Research Laboratories from 1996 to 2006. He has been a Visiting Researcher with ATR Adaptive Communications Research Laboratories since 2006. From 2003 to 2010, he was an Invited Researcher with the NTT Communication Science Laboratories, where he supervised the Open Laboratory for Chaos Information Processing. In 2007, he founded Telecognix Corporation, where he is CEO and head of research. His research interests include analysis and control of complex dynamics for signal generation, processing and communication.