

Bayesian Inference Based Choquet Integral Method

Chunxiao Ren^{1,2}(✉)

¹ Shandong Science and Technology Exchange Center,
No. 607 Shunshua Rd, High-Tech Development Zone, Jinan 250101, China
alanren@163.com

² Shandong Engineering Research Center of Special Optical Fiber and Cable,
No. 1 Optical Fiber and Cable Industrial Park, Yanggu, China

Abstract. The non-additive measure provides a useful tool for many problems in different communities. The Choquet integral has been successfully used for many applications. However, their applicability is restricted due to the exponential computational complexity. In this paper, a novel polynomial method is proposed to solve the parameter estimation problem for Choquet integral. The basic idea of our method is to regard the problem as a sequential one at first; and then we use Bayesian inference method to solve the problem. Using our method, the computational complexity for the non-additive measure is reduced from $O((n+K)*2^{2n})$ to $K*O(n^2 \log n)$. Specifically, the semantic information of the 2^n variables is not lost. This method can be utilized to real-time applications. We provide statistical performance guarantees for the proposed method. As a real-world application, cross-layer design of wireless multimedia networks is optimized by the proposed method. The experiments show that the performance achieved by using this method is better than that of traditional methods.

Keywords: Non-additive measures · Choquet integral
Bayesian inference · Maximum likelihood estimation · Cross-layer

1 Introduction

Within the fields of economics, finance, computer science and decision theory there is an increasing interest in the problem how to replace the additivity property of probability measures by that of monotonicity or, more generally, a non-additive measure. Several types of integrals with respect to non-additive

C. Ren—The author would like to thank Dr. Song Ci, Dr. Haifeng Guo, Dr. Wei An, Dr. Dalei Wu, Dr. Haiyan Luo and Ms. Yuxiao Wu for their helpful comments and valuable suggestions which improved the quality of this paper. This paper is supported by Shandong Outstanding Young Scientists Foundation under Grant No. BS2013DX047 and Shandong Key Research and Development Project under Grant Nos. 2016GGX101042, 2016JMRH0331.

measures, also known as fuzzy measures [30], were developed for different purposes in various works [2, 14, 21, 32, 37]. The Choquet integral, as a popular representation of the non-additive measure, has been successfully used for many applications such as information fusion [7], multiple regressions [33], classification [15, 38], multicriteria decision making [9], image and pattern recognition [20, 39], and data modeling [12].

In general, the basic idea to solve the non-additive model based on Choquet integral is a two-step procedure. The first step is to reduce the non-linear multi-regression model to the traditional linear multi-regression model by converting each n -dimensional vector to a 2^n -dimensional one, which is defined over the powerset of attributes; and thus, the second step is to solve the linear model by using various numerical indices and optimization method [27]. So far, numerous related works have been developed [16, 19, 22, 23, 26].

However, the numbers of variables involved in solving the non-additive model increases exponentially with n and so will the computational time. Thus, the use of non-additive measure in practical applications is clearly curbed by this exponential complexity [24]. Several approaches to deal with this problem are known. The notion of k -additivity proposed by Grabisch [6, 10, 11, 13] enables to find a trade-off between the complexity of representation and the richness of the modeling. In [40], Yan developed a hierarchical Choquet integral to model the data. These techniques, however, are all working on solving the complexity problem through discarding or integrating the 2^n variables to smaller ones. The results obtained through these techniques are all approximate solutions. On the other hand, these results cannot be mapping back to the original 2^n -dimensional space.

To solve the computational complexity problem, several approaches are known as follows. The notion of k -additivity proposed by Grabisch [6, 10, 11, 13] enables to find a trade-off between the complexity of representation and the richness of the modeling. In [40], Yan developed a hierarchical Choquet integral to model the data. These techniques, however, are all working on solving the complexity problem through converting the 2^n -dimensional space to a smaller but less precise one. In the original 2^n -dimensional space, the practical Choquet integral problem is still not solved.

In this paper, we propose a novel polynomial method to solve the parameter estimation problem for Choquet integral. The basic idea of our method is to regard the problem as a sequential one at first; and then we use Bayesian inference method to solve the problem. The parameters of Choquet integral are updated after each data sample presentation. This update procedure is finished in a low-dimensional space mapping from original one. We provide statistical performance guarantees for the proposed method. The experiments show that the performance achieved by using this method is better than that of traditional methods. A special benefit of our method is that the 2^n variables are retained. So the semantic information of the 2^n variables is not lost. Based on our previous works [19, 25, 35, 36], this sequential method is very suitable to real-world applications.

The main contributions of the paper can be summarized as follows. (1) A novel polynomial method is proposed to solve the parameter estimation problem for Choquet integral. Using our method, the computational complexity for the non-additive measure is reduced from $O((n + K) * 2^{2n})$ to $K * O(n^2 \log n)$. (2) As a case of combining sequential concept and Choquet integral, this method can be utilized to real-time applications. (3) As a real-world application, cross-layer design of wireless multimedia networks is optimized by the proposed method.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the preliminaries which provides a mathematical setting for fuzzy measures. Section 3 gives the Bayesian inference based algorithm to estimate the parameters of large-scale Choquet integral. A benchmark dataset test and a real-world application are provided to illustrate the proposed algorithms in Sect. 4. Finally, summary and future works are drawn in Sect. 5.

2 Choquet Integral

2.1 Basic Definitions

Let us consider a multi-feature problem described by a finite set of alternatives $A = \{a_1, a_2, \dots\}$ and a finite set of features $F = \{f_1, f_2, \dots\}$. Each alternative $a_i \in A$ can be associated with a profile $(\mu_1, \dots, \mu_n) \in \mathcal{R}^n$, where, for any $f_j \in F, \mu_j \in \mathcal{R}$ represents the utility of a_i related to feature j .

In general, one can compute an aggregation score $H(f, w)$ by a linear basis function model which takes into account the weights of importance of the feature.

$$H(f, w) = \sum_{j=1}^n w_j \phi_j(f) = w^T \phi(f) \tag{1}$$

where $w = (w_1, \dots, w_n)^T$ and $\phi = (\phi_1, \dots, \phi_n)^T$. The weight w is also regarded as a Lebesgue measure on a singleton f . However, the assumption of independence among features is rarely verified. In order to take into account a flexible representation of complex interaction phenomena among features, it is useful to substitute to the weight vector w a non-additive set function μ on $N = 1, \dots, n$, called Choquet capacity [34], allowing defining a weight not only on the importance of each feature, but also on the importance of each subset of features. It is clearly more powerful than the Lebesgue integral model since Lebesgue integral thus becomes a special case of the Choquet integral model. The Choquet integral is defined as follows [4, 7].

Definition 1. For every space Ω and algebra \mathcal{A} of subset of Ω , a set-function $\mu : \mathcal{A} \rightarrow \mathcal{R}$ is called a capacity if it satisfies the following:

- (i) $\mu(\emptyset) = 0, \mu(\Omega) = 1,$
- (ii) $\forall A, B \in \mathcal{A} : A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

Furthermore, a capacity μ on Ω is said to be additive if

- (iii) $\forall A, B \in \mathcal{A} : \mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$

For each subset of feature $A \subseteq \Omega$, the number $\mu(A)$ can be regarded as the weight or the importance of A .

Definition 2. If $\phi : \Omega \rightarrow \mathcal{R}$ is a bounded A -measurable function and μ is any capacity on Ω , we define the Choquet integral of ϕ with respect to μ to be the number

$$\int_{\Omega} \phi(f) d\mu(f) = \int_0^{\infty} \mu(\{f \in \Omega : \phi(f) \geq \alpha\}) d\alpha + \int_{-\infty}^0 [\mu(\{f \in \Omega : \phi(f) \geq \alpha\}) - 1] d\alpha \tag{2}$$

where the integrals are taken in the sense of Riemann. In particular, if Ω is finite and $\phi(f_1) \geq \phi(f_2) \geq \dots \geq \phi(f_n)$, then

$$\int_{\Omega} \phi(f) d\mu(f) = \sum_{i=1}^{n-1} (\phi(f_i) - \phi(f_{i+1})) \mu(\{f_1, \dots, f_i\}) + \phi(f_n) \tag{3}$$

Definition 3. Let ϕ be a function from $X = \{x_1, \dots, x_n\}$ to $[0, 1]$. Let $\{x_{\sigma(1)}, \dots, x_{\sigma(n)}\}$ denote a reordering of the set X such that $0 \leq \phi(x_{\sigma(1)}) \leq \dots \leq \phi(x_{\sigma(n)})$, and let $A_{(i)}$ be a collection of subsets defined by $A_{(i)} = \{x_{\sigma(i)}, \dots, x_{\sigma(n)}\}$. Then, the discrete Choquet integral of ϕ with respect to a fuzzy measure μ on X is defined as

$$C_{\mu}(\phi) = \sum_{i=1}^n \mu(A_{(i)}) (\phi(x_{(i)}) - \phi(x_{(i-1)})) = \sum_{i=1}^n \phi(x_{(i)}) (\mu(A_{(i)}) - \mu(A_{(i+1)})) \tag{4}$$

where we take $\phi(x_{(0)}) = 0$, $A_{(n+1)} = 0$, and $x_{(i)} = x_{\sigma(i)}$.

2.2 Existing Method

Based on the Definitions 1–3, the relationship between the Choquet integral of ϕ and the capacity μ can also be described by a new nonlinear multi-feature regression model [5]:

$$C_{\mu}(\phi) = e + \int_{\Omega} \phi(f) d\mu(f) + \mathcal{N}(0, \delta^2) \tag{5}$$

where e is a regression constant, ϕ is an observation of $X = \{x_1, x_2, \dots, x_n\}$, $N(0, \delta^2)$ is a normally distributed random perturbation with expectation 0 and variance δ^2 , and δ^2 is the regression residual error. The Choquet integral problem is reduced to a traditional linear multi-regression model. At the

same time, specifying a general fuzzy measure requires specification of $2^n - 1$ parameters, which is clearly exponential.

For solving this linear multi-regression model, usually we need to define a loss function. Through minimizing this loss function, the parameter of the linear model can be derived. Given classes C_1, \dots, C_n , Grabisch [17, 18] proposed a mean squared error (MSE) criterion, where the difference between desired outputs t_i for $i = 1, \dots, n$ and the actual outputs $C_\mu(\phi)$ is minimized under constraints [1]. The loss function L is

$$L^2 = \sum_{\phi \in C_1} (C_\mu(\phi) - t_1)^2 + \dots + \sum_{\phi \in C_n} (C_\mu(\phi) - t_n)^2 \tag{6}$$

Given the observation data, the optimal regression coefficients μ can be determined by using regression methods. For example, in order to make the loss function minimal, the least square method, as a regression method, is used for determine $\mu_k (k = 1, 2, \dots, 2^n - 1)$ [28].

We use a maximum likelihood estimation (MLE) [8] method to fulfill the least square method. In fact, we define the likelihood function by using Gaussian distribution:

$$p(t|\phi, \mu, \beta) = \mathcal{N}(t|C_\mu(\phi), \beta^{-1}) \tag{7}$$

where t is desired outputs given by a deterministic function $C_\mu(\phi)$, and β is the precision of Gaussian random variable. Consider an observation set of inputs ϕ with corresponding desired outputs t_1, \dots, t_n ,

$$p(t|\phi, \mu, \beta) = \prod_{n=1}^N \mathcal{N}(t_n|\mu^T \phi_n, \beta^{-1}) \tag{8}$$

where t and μ are column vectors

Solving for μ using MLE method, we obtain

$$\mu_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t \tag{9}$$

Here Φ is an $N \times M$ matrix whose elements are given by $\Phi_{nj} = \phi_j(f_n)$ [31], so that

$$\Phi = \begin{pmatrix} \phi_0(f_1) & \phi_1(f_1) & \dots & \phi_{M-1}(f_1) \\ \phi_0(f_2) & \phi_1(f_2) & \dots & \phi_{M-1}(f_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(f_N) & \phi_1(f_N) & \dots & \phi_{M-1}(f_N) \end{pmatrix} \tag{10}$$

To summarize, the algorithm to be employed by MLE method is as follows:

For time-complexity analysis, we assume that there are n feature and K samples. The Choquet integral space is 2^n -dimensional. In addition, we assume that the complexity of the inversion of a matrix $n \times n$ is $\Omega(n^2 \log n)$ [29]. We present the time complexity using pseudo-code analysis.

Algorithm 1. MLE-based Choquet integral

Input: matrix Φ , target variable t

Output: μ

- Step 1: Compute the $\Phi^T \Phi$
 - Step 2: Compute the $\Phi^T t$
 - Step 3: Compute the $(\Phi^T \Phi)^{-1}$
 - Step 4: Compute the $(\Phi^T \Phi)^{-1} \Phi^T t$
-

$O(K(2^n)^2)$	for $\Phi^T \Phi$
$O(K * 2^n)$	for $\Phi^T t$
$\Omega((2^n)^2 \log 2^n)$	for $(\Phi^T \Phi)^{-1}$
$O((2^n)^2)$	for $(\Phi^T \Phi)^{-1} \Phi^T t$

We can rewrite this time complexity as

$$O(K * 2^{2n} + K * 2^n + \log 2 * n * 2^{2n} + 2^{2n})$$

Thus, we have that the time complexity for MLE-based method is $O((n+K) * 2^{2n})$. Obviously, the numbers of variables involved in solving the non-additive model increases exponentially with n and so will the computational time. In fact, the use of non-additive measure in practical applications is clearly curbed by this exponential complexity [24]. As an important challenge, the practical application of the non-additive measure has been plagued by real-time computational problems for a long time.

On the other hand, the traditional method, such as the MLE-based method, which involves processing the entire samples in one go, can be inappropriate for some real-time applications in which the data observations are arriving in a continuous stream, and predictions must be made before all of the samples are acquired. For a large scale dataset, it may be suitable to use sequential algorithms in which the data samples are considered one by one, and the model parameters updated after each such presentation.

3 Our Method

A general Choquet integral is defined by $2^n - 1$ coefficients. In order to reduce the computational complexity, we developed a Bayesian inference based Choquet integral (BIBCI) method.

Now we introduce some theorems needed in this work, and these theorems also can be found in [3].

Given a marginal Gaussian distribution for ϕ and a conditional Gaussian distribution for $C_\mu(\phi)$ given ϕ in the form

$$p(\phi) = \mathcal{N}(\phi | \varphi, P^{-1}) \tag{11}$$

$$p(C_\mu(\phi) | \phi) = \mathcal{N}(C_\mu(\phi) | \mu\phi + b, Q^{-1}) \tag{12}$$

where $\varphi, \mu,$ and b are parameters governing the means, and P and Q are precision matrices. So the marginal distribution of $C_\mu(\phi)$ and the conditional distribution of ϕ given $C_\mu(\phi)$ are given by

$$p(C_\mu(\phi)) = \mathcal{N}(C_\mu(\phi) | \mu\varphi + b, Q^{-1} + \mu P^{-1} \mu^T) \tag{13}$$

$$p(\phi | C_\mu(\phi)) = \mathcal{N}\left(\phi \left| \sum \{ \mu^T Q (C_\mu(\phi) - b) + P\varphi \}, \Sigma \right.\right) \tag{14}$$

where

$$\Sigma = (P + \mu^T Q \mu)^{-1} \tag{15}$$

Based on (3)–(7), we can derive the posterior distribution of μ as

$$p(\mu | t) = \mathcal{N}(\mu | \varphi_N, S_N) \tag{16}$$

where

$$\varphi_N = S_N (S_0^{-1} \varphi_0 + \beta \Phi^T t) \tag{17}$$

$$S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi \tag{18}$$

where φ_N is prior means, S_N is prior precision matrices, β is prior uncertainty. Because the posterior distribution is Gaussian, its mode coincides with its mean. Thus the maximum posterior weight vector is simply given by $\mu_{MAP} = \varphi_N$.

Suppose that we have already observed N data points; thereby the posterior distribution over μ is given. This posterior can be regarded as the prior for the next observation. By considering an additional data point (ϕ_{N+1}, t_{N+1}) the resulting posterior distribution can be derived based on (8)–(10)

$$p(\mu | t_{N+1}, \phi_{N+1}, \varphi_N, S_N) = \mathcal{N}(\mu | \varphi_{N+1}, S_{N+1}) \tag{19}$$

with

$$S_{N+1}^{-1} = S_N^{-1} + \beta \phi_{N+1} \phi_{N+1}^T \tag{20}$$

$$\varphi_{N+1} = S_{N+1} (S_N^{-1} \varphi_N + \beta \phi_{N+1} t_{N+1}) \tag{21}$$

Since the non-additive measure in the Choquet model is defined over the powerset Ω , the reduction step basically aggregates the observed data of individual features to the observation on sets. It is clear that there are only $n(n \ll 2^n)$ non-zero elements in each 2^n -dimensional vector data. On the other hand, it is not all necessary to use the 2^n coefficients for building a Choquet integral model. In most cases, we only need a very small part of these coefficients to compute the Choquet integral. More importantly, some coefficients may not be used at all in many applications. However, for traditional Choquet integral method, the model cannot be built if we do not figure out all the coefficients.

We now develop a new approximation algorithm to reduce the computational complexity. Because most information about the relationship among features is

contained in only a small fraction of all the coefficients, we can use the n non-zero element in each sample to adjust the related parameter. For each sample, the parameters of Choquet integral related with non-zero element are updated through Bayesian inference method. The basic idea behind the proposed method is that it is far easier to consider a sequence of conditional distributions than it is to obtain the marginal by integration of the joint density. Based on (19)–(21), we can update the parameters of Choquet integral model by Bayesian inference. Though only n parameters are modified in each training process, the 2^n -dimensional parameter vector will tend to be reasonable after certain rounds of training. Meanwhile, this algorithm only updates the coefficients that the training data support. It is possible to acquire a reasonable model which maybe has a few coefficients much less than 2^n . It then executes the following algorithm.

We present the time complexity using pseudo-code analysis. We can rewrite this time complexity as

$$K * O(n + n^2 + n^2 + n^2 \log n + n^2 + n + n + n + n)$$

Thus, the time complexity for our method is $K * O(n^2 \log n)$.

Our method uses the idea of mapping high-dimensional space distributions to lower one to update the utility parameters.

Given an n -dimensional vector in original non-additive space, it can be mapped to a 2^n -dimensional linear space. If the our method use K samples to achieve convergence, and we define a learning rate as λ_p under certain precision p , the frequency of training of each element in an n -dimensional vector is generally in a direct ratio to n :

$$\frac{nK}{2^n} \approx \lambda_p n \tag{22}$$

Algorithm 2. Bayesian Inference Based Choquet Integral Method

Input: matrix Φ , target variable t

Output: μ

- Initialize model parameters $\{\beta, \mu_0\}$
- For $k=1, \dots, N$
 1. Sample the data ϕ_k and target variable t_k .
 2. Obtain the n -dimensional non-zero vector $\phi_k^{(n)}, \mu_{k-1}^{(n)}$ and $t_k^{(n)}$
 3. Initialize n -dimensional prior precision matrices $\{S^{(n)}\}$
 4. Compute the posterior precision matrices $S_k^{(n)}$:

$$\left(S_k^{(n)}\right)^{-1} = \left(S^{(n)}\right)^{-1} + \beta \phi_k^{(n)} \left(\phi_k^{(n)}\right)^T$$

- 5. Compute the posterior $\mu_k^{(n)}$:

$$\mu_k^{(n)} = S_k^{(n)} \left(\left(S^{(n)}\right)^{-1} \mu_{k-1}^{(n)} + \beta \phi_k^{(n)} t_k^{(n)} \right)$$

- 6. Update the n related parameters back to μ_k

Usually, our method needs K samples to achieve the precision p , and K can be estimated by:

$$K \approx 2^n \lambda_p \quad (23)$$

4 An Application Example

For wireless multimedia networks, cross-layer design has been regarded as one of the most effective and efficient ways to provide quality of service (QoS). At the application layer, prediction mode and quantization parameter (QP) in video encoding are two critical design variables [H.264 2005]. At the physical layer, modulation and coding schemes (MCS) have been adopted to achieve a good tradeoff between transmission rate and transmission reliability. PSNR indicates the performance of wireless multimedia networks. The channel signal-to-interference-noise ratio (SINR) reflects the channel conditions

Using non-additive measure, we can estimate the parameters $\mu_k (k = 1, 2, \dots, n)$. In fact, μ_k indicates the importance of coefficients k . In previous works [19, 25, 35, 36], we have illustrated that the parameters of model is stable under certain scenario, and the optimization using non-additive measure for cross-layer design is effective when channel condition is known.

Let us further consider this issue for real-world applications. In fact, it is very hard to acquire the real-time channel condition information in wireless multimedia networks community. In other words, the system is usually unaware of the state transitions. For example, we can only obtain the coefficients of QP and MCS for a wireless multimedia networks system. The optimal configuration of QP and MCS is largely determined by SINR under the current scenario. However, the system is not able to obtain the current SINR.

For cross-layer design optimization of wireless multimedia networks, the three problems mentioned above all exist. The traditional non-additive measure method is not able to handle this multi-modal, real-time and high complexity problem.

4.1 The Performance of the Application

The MLE and our methods were applied to the real-time transmission of an individual video bitstream across a multi-hop 802.11a/e wireless network. We first discuss the regression experiments and then the real-time application experiments.

Regression. In this part, the data set contained 8064 3-D samples, each containing one PSNR value from each of the three variables (Mac_length, QP and AMC) used in the cross-layer optimization problem. Two algorithms, MLE and Bayesian inference based Choquet integral (BIBCI) methods, were considered for non-additive measure.

To evaluate the methods as regression algorithms, three runs of a 10-fold cross-validation (in a total of 30 simulations) were applied for the MLE method. As real-time methods, our methods were applied directly to real-time prediction. The overall performance is evaluated by MAD and RMSE. We also conducted the paired t-tests against true PSNR value confirmed the statistical significance of each method.

The results are shown in Table 1. Under the MAD and RMSE criterion, the BIBCI method outperforms the MLE method. For the time consumption, the MLE method is better than BIBCI method. This is an interesting outcome, since the matrix of variables can be calculated directly under low-dimensional situation, with no need for accumulated calculations. However, from the MAD and RMSE points of view, the best option is still our method.

Table 1. The experiment results in terms of the MAD errors and t-test

	MAD	RMSE	Mean	t	p_{one}	p_{two}	Time
MLE(1)	4.935	6.775	29.35	-0.008	0.497	0.994	0.00042
MLE(2)	4.936	6.776	29.35	0.0004	0.5	1.0	0.00023
MLE(3)	4.931	6.769	29.44	0.00004	0.5	1.0	0.00039
BIBCI	1.788	2.694	29.34	0.09	0.497	0.93	6.71

Real-Time Application. In this part, we simulate the real time wireless multimedia dataset. In this dataset, channel condition is unstable and constantly changing. We draw samples to simulate the real-life situation where the channel condition changes from bad to good or vice versa. There are 1300 samples in this dataset, and it includes 2 period of channel condition changing.

For better illustration, we process the dataset using kernel smoothing as shown in Figs. 1 and 2. The channel condition is just utilized to acquire the optimal PSNR. In real-life situation, the channel condition is hidden and unknown.

For the MLE method, the utility of variable can be calculated by traversing all the options. For our dataset, the utility of Mac_length equals to 28.56, the utility of QP equals to 35.73 and the utility of AMC equals to 38.40. So the distortion performance is more sensitive to the AMC than to other variables. In this test, system optimizes the AMC configuration all the time for the MLE method.

For our method, we calculate the utilities for each slice, and consider the variable with the largest utility value as the most important variable. For each slice, system always optimizes the most important variable. In order to avoid local optimum, we update the utilities by 8 random samples before dealing with each slice.

The results are shown in Fig. 3. We can see clearly that the performance of our method is closer to the optimal than traditional Choquet integral. We also can find that the accuracy of the first 200 samples is not very satisfactory,

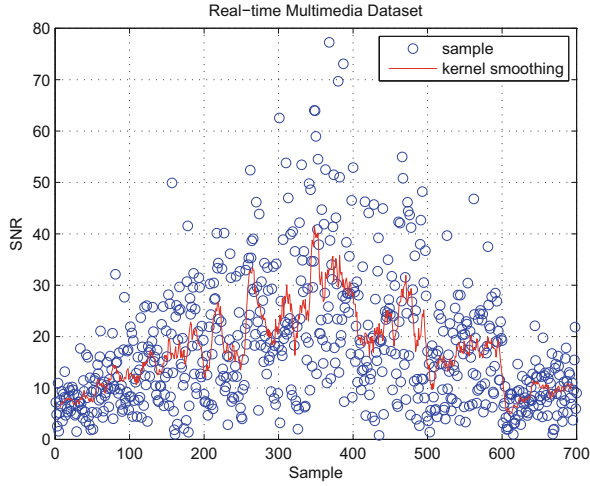


Fig. 1. The kernel smoothing of the real-time wireless multimedia dataset

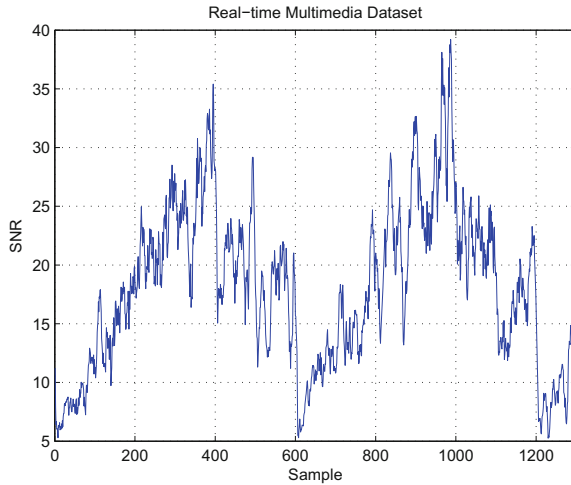


Fig. 2. The SNR of the real-time wireless multimedia dataset

and this phenomenon does not happen in second period, which means that our algorithm turns to stable within 9×200 updates under this dataset.

The performances of MLE and our methods are shown in Table 2. Larger values result in better multimedia system. To illustrate the quality of the reconstructed videos, we show some sample video frames dealing with the optimal configuration, the MLE-based method and our method, respectively. As shown in Table 3, we can see that the performance achieved by using our method is better than that of MLE-based method.

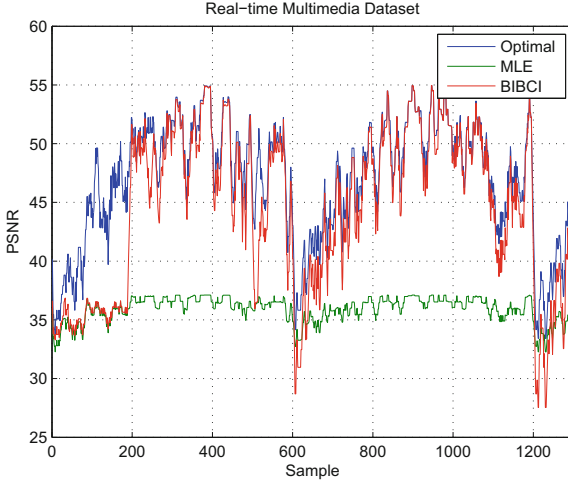
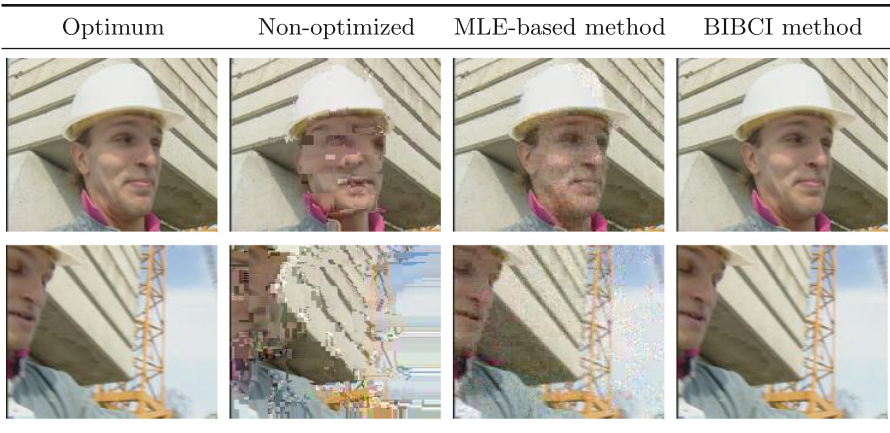


Fig. 3. The simulation on the real-life situation. The performance of our method is closer to the optimal than traditional Choquet integral.

Table 2. The mean performance of MLE and BIBCI methods

	Optimum	MLE	BIBCI
Mean PSNR	46.05	35.77	42.71

Table 3. Some sample video frames dealing with the optimal configuration, the MLE-based method and BIBCI method.



5 Conclusion

In this paper, we propose a novel polynomial method to solve the parameter estimation problem for Choquet integral. The proposed approach allows training parameters of Choquet integrals without requiring entire input samples. The computational complexity of the proposed algorithm is low and the number of parameters needed to compute is only linear rather than exponential with the number of inputs. Using our method, the computational complexity for the non-additive measure is reduced from $O((n+K)*2^{2n})$ to $K*O(n^2 \log n)$. Specifically, the semantic information of the 2^n variables is not lost. This method can be utilized to real-time applications. As a real-world application, cross-layer design of wireless multimedia networks is optimized by the proposed method. The experiments show that the performance achieved by using this method is better than that of traditional methods.

References

1. Andres, M., Paul, G.: Minimum classification error training for choquet integrals with applications to landmine detection. *IEEE Trans. Fuzzy Syst.* **16**(1), 225–238 (2008)
2. Aumann, R., Shapley, L., Aumann, R.: *Values of Non-atomic Games*. Princeton University Press, Princeton (1974)
3. Bishop, C., et al.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
4. Choquet, G.: Theory of capacities. *Ann. Inst. Fourier* **5**(131–295), 54 (1953)
5. Ci, S., Guo, H.: Quantitative dynamic interdependency measure and significance analysis for cross-layer design under uncertainty. In: *Proceedings of 16th International Conference on Computer Communications and Networks, ICCCN 2007*, pp. 900–904 (2007)
6. Fujimoto, K.: New characterizations of k-additivity and k-monotonicity of bi-capacities. In: *2nd International Conference on Soft Computing and Intelligent Systems and 5th International Symposium on Advanced Intelligent Systems, SCISIS 2004* (2004)
7. Gilboa, I., Schmeidler, D.: Additive representations of non-additive measures and the Choquet integral. *Ann. Oper. Res.* **52**(1), 43–65 (1994)
8. Golub, G., Van Loan, C.: *Matrix Computations*. Johns Hopkins University Press, Baltimore (1996)
9. Grabisch, M.: Fuzzy integral in multicriteria decision making. *Fuzzy Sets Syst.* **69**(3), 279–298 (1995)
10. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Syst.* **92**(2), 167–189 (1997)
11. Grabisch, M.: The interaction and Mobius representations of fuzzy measures on finite spaces, k-additive measures: a survey. In: *Fuzzy Measures and Integrals Theory and Applications*, pp. 70–93 (2000)
12. Grabisch, M.: Modelling data by the Choquet integral. *Stud. Fuzziness Soft Comput.* **123**, 135–148 (2003)
13. Grabisch, M., Labreuche, C.: Bi-capacities-I: definition, Mobius transform and interaction. *Fuzzy Sets Syst.* **151**(2), 211–236 (2005)

14. Grabisch, M., Nguyen, H., Walker, E.: *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Springer, Dordrecht (1995). <https://doi.org/10.1007/978-94-015-8449-4>
15. Grabisch, M., Nicolas, J.: Classification by fuzzy integral: performance and tests. *Fuzzy Sets Syst.* **65**(2–3), 255–271 (1994)
16. Grabisch, M., Raufaste, E.: An empirical study of statistical properties of the Choquet and Sugeno integrals. *IEEE Trans. Fuzzy Syst.* **16**(4), 839 (2008)
17. Grabisch, M., Sugeno, M.: Multi-attribute classification using fuzzy integral. In: *IEEE International Conference on Fuzzy Systems 1992*, pp. 47–54 (1992)
18. Grabisch, M., Sugeno, M., Murofushi, T.: *Fuzzy Measures and Integrals: Theory and Applications*. Springer-Verlag, New York Inc., Secaucus (2000)
19. Guo, H., Zheng, W., Ci, S.: Precise determination of non-additive measures. In: *2008 IEEE Conference on Cybernetics and Intelligent Systems*, pp. 911–916 (2008)
20. Keller, J., Gader, P., Hocaoglu, A.: Fuzzy integrals in image processing and recognition. In: *Fuzzy Measures and Integrals: Theory and Applications*, pp. 435–466 (2000)
21. Klement, E., Mesiar, R., Pap, E.: A universal integral as common frame for Choquet and Sugeno integral. *IEEE Trans. Fuzzy Syst.* **18**(1), 178–187 (2010)
22. Kojadinovic, I.: Minimum variance capacity identification. *Eur. J. Oper. Res.* **177**(1), 498–514 (2007)
23. Kojadinovic, I.: Quadratic distances for capacity and bi-capacity approximation and identification. *4OR: Q. J. Oper. Res.* **5**(2), 117–142 (2007)
24. Kojadinovic, I., Labreuche, C.: Partially bipolar Choquet integrals. *IEEE Trans. Fuzzy Syst.* **17**(4), 839–850 (2009)
25. Luo, H., Argyriou, A., Wu, D., Ci, S.: Joint source coding and network-supported distributed error control for video streaming in wireless multihop networks. *IEEE Trans. Multimed.* **11**(7), 1362–1372 (2009)
26. Marichal, J.: An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Trans. Fuzzy Syst.* **8**(6), 800–807 (2000)
27. Miranda, P., Grabisch, M.: Optimization issues for fuzzy measures. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **7**(6), 545–560 (1999)
28. Rao, C., Mitra, S.: *Generalized Inverse of the Matrix and Its Applications*. Wiley, New York (1971)
29. Rota, G.: On the foundations of combinatorial theory I. Theory of Möbius functions. *Probab. Theory Relat. Fields* **2**(4), 340–368 (1964)
30. Sugeno, M.: *Theory of fuzzy integrals and its applications*. Tokyo Institute of Technology (1974)
31. Tveit, A.: On the complexity of matrix inversion. *Mathematical Note, IDI, NTNU, Trondheim, Norway* (2003)
32. Verma, N., Hanmandlu, M.: Additive and nonadditive fuzzy hidden Markov models. *IEEE Trans. Fuzzy Syst.* **18**(1), 40–56 (2010)
33. Wang, Z., Guo, H.: A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals. In: *Proceedings of the FUZZIEEE 2003*, pp. 819–821 (2003)
34. Wang, Z., Leung, K., Klir, G.: Applying fuzzy measures and nonlinear integrals in data mining. *Fuzzy Sets Syst.* **156**(3), 371–380 (2005)
35. Wu, D., Ci, S., Wang, H.: Cross-layer optimization for video summary transmission over wireless networks. *IEEE J. Sel. Areas Commun.* **25**(4), 841–850 (2007)
36. Wu, D., Ci, S., Luo, H., Guo, H.: A theoretical framework for interaction measure and sensitivity analysis in cross-layer design. *ACM Trans. Model. Comput. Simul.* (2010)

37. Wu, W., Leung, Y., Mi, J.: On generalized fuzzy belief functions in infinite spaces. *IEEE Trans. Fuzzy Syst.* **17**(2), 385–397 (2009)
38. Xu, K., Wang, Z., Heng, P., Leung, K.: Classification by nonlinear integral projections. *IEEE Trans. Fuzzy Syst.* **11**(2), 187–201 (2003)
39. Xu, K., Wang, Z., Wong, M., Leung, K.: Discover dependency pattern among attributes by using a new type of nonlinear multiregression. *Int. J. Intell. Syst.* **16**(8), 949–962 (2001)
40. Yan, N., Wang, Z., Chen, Z.: Classification with Choquet integral with respect to signed non-additive measure. In: *ICDMW*, pp. 283–288 (2007)