Regular paper | Journal of the KIECS. pp. 1023-1030, vol. 15, no. 6, Dec. 31. 2020, t. 104, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2020.15.6.1023

임의의 그래프신호를 위한 고속 샘플링 집합 선택 알고리즘

김윤학*

Fast Sampling Set Selection Algorithm for Arbitrary Graph Signals

Yoon-Hak Kim*

요 약

임의의 그래프 신호를 복원하기 위해 그래프상의 일부 노드로 구성된 샘플링 집합내의 노드들의 신호값만을 사용하게 되는 경우, 이를 위한 최적의 샘플링 집합 선택 문제에 대해 연구한다. 고도의 계산량을 요구하는 고유값 분해 (eigen decomposition)를 사용하지 않고, 노드를 선택하는 과정에서의 신호 변화값의 차이를 비용함수로 제시한다. 구체적으로, 기존 방식의 비용함수인 신호 복원오차를 최소화하는 대신에 본 연구에서는 신호 변화값의 차이를 비용함수로 채택하여 이를 최소화하는 간단하고 고속의 탐욕 (greedy) 샘플링 집합 선택 알고리즘을 제안한다. 기존의 고속알고리즘과 성능평가 비교를 위해 다양한 그래프 신호에 대한 폭넓은 실험을 진행하여, 기존 방식 대비 신호복원 성능감소를 약 7% 이내로 유지하면서 실행시간을 10배이상으로 단축하였음을 보인다.

ABSTRACT

We address the sampling set selection problem for arbitrary graph signals such that the original graph signal is reconstructed from the signal values on the nodes in the sampling set. We introduce a variation difference as a new indirect metric that measures the error of signal variations caused by sampling process without resorting to the eigen-decomposition which requires a huge computational cost. Instead of directly minimizing the reconstruction error, we propose a simple and fast greedy selection algorithm that minimizes the variation differences at each iteration and justify the proposed reasoning by showing that the principle used in the proposed process is similar to that in the previous novel technique. We run experiments to show that the proposed method yields a competitive reconstruction performance with a substantially reduced complexity for various graphs as compared with the previous selection methods.

키워드

Graph Signal Processing, Sampling Set Selection, Greedy Algorithm, Signal Variation, Signal Reconstruction
그래프 신호 처리, 샘플링 집합 선택, 탐욕 알고리즘, 신호 변화, 신호 복원

* 교신저자 : 조선대학교 전자공학부
• 접 수 일 : 2020. 09. 28
• 수정완료일 : 2020. 11. 07
• 게재확정일 : 2020. 12. 15
• Received : Sep. 28, 2020, Revised : Nov. 07, 2020, Accepted : Dec. 15, 2020
• Corresponding Author : Yoon-Hak Kim
Dept. Electronic Engineering, Chosun University,
Email : yhk@chosun.ac.kr

Ⅰ. Introduction

Recently, a high volume of data samples are encountered in network applications such as neural, transportation, energy and social and sensor networks. In contrast to traditional data, these reside on irregularly structured networks. Graphs enable us to represent such high-dimensional data by using the vertices of graphs and the edges with weights. Specifically, graph signals are defined on the vertices or nodes of graphs to represent data samples on nodes and signal smoothness is described from the weight of edges. For instance, since two nodes close to each other are likely to generate similar data samples, the edge connecting them may take a weight inversely proportional to the physical distance [1, 2]. To measure signal variations caused by irregular structure of graphs, variation operators such as the combinatorial graph Laplacian, and the normalized Laplacian can be introduced [2, 4].

Graph signal processing handles such network data to achieve certain objectives. Since the data is generally high-dimensional, the techniques to reduce the size of data samples have been developed in recent years [3-10]. The goal of the techniques is to select a subset of vertices of the graph such that the original signal can be recovered from data on the sampling set. To avoid a huge computational cost in finding the best sampling set, greedy selection methods have been developed in [3-5] to minimize the worst case of the reconstruction error. To find the second-order statistics of graph signals, a greedy sampling method was presented in [6]. A near-optimality of greedy algorithm was shown by using the concept of approximate submodularity [7]. Nothing that data samples can be easily inferred from the nodes densely connected, a non-uniform sampling based on the local uncertainty principle which selects more samples in area of high concentration was devised [8]. Since most of the sampling set selection methods require the eigen-decomposition of matrices, algorithms without requiring such decomposition was presented to achieve a fast execution time and reasonable performance [9, 10].

In this work, we consider the sampling set selection without the eigen-decomposition to facilitate fast selection process in practical applications. We focus on the signal variation and aim to take a greedy selection of nodes of the graph such that the variation difference is minimized at each iteration. In contrast to most of the previous methods which run on the eigenvector matrix of variation operators, we directly employ the variation operator to devise a fast and low-weight algorithm. We also discuss that the proposed method works based on the principle similar to that of the non-uniform sampling presented in [8]. We examine the performance of the proposed algorithm by experiments and show that our algorithm produces a competitive reconstruction performance with a fast execution time at the cost of a substantially reduced complexity as compared with previous sampling selection methods.

The main contributions of this work are: (a) we propose a fast sampling set selection that directly works on variation operators to avoid a heavy computation of the eigen-decomposition of the operators, leading to a low-weight selection process and (b) we do not restrict to bandlimited graph signals which have been assumed for many previous methods to simplify their derivation processes. Thereby, our proposed algorithm is expected to perform well for arbitrary graph signals.

This paper is organized as follows. The problem is formulated in Section II. The proposed algorithm is explained and summarized in Section III. Comparison with the previous work [8] is provided to justify the principle applied to the proposed

algorithm in Section IV. The performance of the proposed algorithm is demonstrated by experiments for various graphs in Section V and the conclusion given in Section VI.

II. Problem Formulation

We consider a graph $G(V,E)$ with N vertices denoted by a set of nodes $V=\{1,\cdots,N\}$ and edges $E=\{(i,j,w_{ij})\}$ where w_{ij} is a weight of the edge associated with node i and j. A set of N data samples on V is regarded as a graph signal $\boldsymbol{f} \in \mathbf{R}^N$ with signal value f_i indicating the sample on the i-th vertex. From the connectivity of nodes of graphs, variation operators (e.g., combinatorial graph Laplacian, normalized Laplacian) can be produced to evaluate the variation of signals over nodes of graphs [2, 4]. Assuming that the variation operator $\boldsymbol{L}$, $N\times N$ matrix, has orthonormal eigenvectors $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_N$ with corresponding eigenvalues $|\lambda_1|\leq\cdots\leq|\lambda_N|$, the graph signal $\boldsymbol{f}$ can be written by

$$\boldsymbol{f}=\boldsymbol{U}\tilde{\boldsymbol{f}}=\sum_{i=1}^{N}\tilde{f}(\lambda_i)\boldsymbol{u}_i \tag{1}$$

where $\boldsymbol{U}=[\boldsymbol{u}_1\cdots\boldsymbol{u}_N]$ is the eigenvector matrix and $\tilde{\boldsymbol{f}}=\boldsymbol{U}^{-1}\boldsymbol{f}=\boldsymbol{U}^T\boldsymbol{f}$ the graph Fourier transform (GFT) of $\boldsymbol{f}$ with entries $\tilde{f}(\lambda_i)$.

Now, we sample the graph signal $\boldsymbol{f}$ to obtain the sampled signal $\boldsymbol{f}_S$ with signal values indexed by a sampling set S. Let $\boldsymbol{S}$ be an $N\times|S|$ matrix consisting of columns, each of which indicates nodes in the set S. Then, the sampled signal can be expressed from (1) as follows:

$$\boldsymbol{f}_S=\boldsymbol{S}^T\boldsymbol{f}=\boldsymbol{S}^T\boldsymbol{U}\tilde{\boldsymbol{f}}=\boldsymbol{U}_{SV}\tilde{\boldsymbol{f}} \tag{2}$$

where $\boldsymbol{U}_{SV}$ is a submatrix of the eigenvector matrix $\boldsymbol{U}$ with rows indexed by S and columns by V, respectively. Assuming that $\boldsymbol{U}_{SV}$ has a rank $|S|$, the least square estimate (LSE) of $\tilde{\boldsymbol{f}}$ is given by $\hat{\tilde{\boldsymbol{f}}}=\boldsymbol{U}_{SV}^{+}\boldsymbol{f}_S$ where $\boldsymbol{U}_{SV}^{+}=U_{SV}^T(\boldsymbol{U}_{SV}\boldsymbol{U}_{SV}^T)^{-1}$ is the pseudoinverse of $\boldsymbol{U}_{SV}$. Thus, the graph signal can be recovered from the sampled one by using the LSE and the reconstructed signal $\hat{\boldsymbol{f}}$ can be given by:

$$\hat{\boldsymbol{f}}=\boldsymbol{U}\hat{\tilde{\boldsymbol{f}}}=\boldsymbol{U}\boldsymbol{U}_{SV}^T(\boldsymbol{U}_{SV}\boldsymbol{U}_{SV}^T)^{-1}\boldsymbol{f}_S \tag{3}$$

Note that most of the previous work focused on finding the optimal sampling set by minimizing the reconstruction error $E\|\boldsymbol{f}-\hat{\boldsymbol{f}}\|^2$ [3-5]. In this work, we seek to minimize a simple and indirect metric in selection process which is introduced in the following section.

III. Fast sampling set selection

Let $\bar{\boldsymbol{f}}_S$ be the sampled signal defined as

$$\bar{\boldsymbol{f}}_S=\begin{cases}f_i & i\in S\\ 0 & i\in S^C\equiv V-S\end{cases} \tag{4}$$

Then, we have $\bar{\boldsymbol{f}}_S=\boldsymbol{I}_S\boldsymbol{f}$ where $\boldsymbol{I}_S$ is a diagonal matrix with entries $d_i=1, i\in S$ and $d_i=0, i\in S^C$. Note that $\boldsymbol{f}^T\boldsymbol{L}\boldsymbol{f}$ represents the variation of signal $\boldsymbol{f}$ over nodes of a graph G. In this work, we aim to find the best sampling set S^* that minimizes the variation difference due to the sampling process: more specifically, for arbitrary graph signal $\boldsymbol{f}$,

$$S^*=\arg\min_S\left|\boldsymbol{f}^T\boldsymbol{L}\boldsymbol{f}-\bar{\boldsymbol{f}}_S^T\boldsymbol{L}\bar{\boldsymbol{f}}_S\right| \tag{5}$$

$$=\arg\min_S\left|\boldsymbol{f}^T\boldsymbol{L}\boldsymbol{f}-\boldsymbol{f}^T\boldsymbol{I}_S^T\boldsymbol{L}\boldsymbol{I}_S\boldsymbol{f}\right| \tag{6}$$

$$=\arg\min_S\left|\boldsymbol{f}^T\left(\boldsymbol{L}-\boldsymbol{I}_S^T\boldsymbol{L}\boldsymbol{I}_S\right)\boldsymbol{f}\right| \tag{7}$$

$$=\arg\min_S\left|\boldsymbol{f}^T\left(\boldsymbol{L}-\bar{\boldsymbol{L}}_S\right)\boldsymbol{f}\right|$$

where $\overline{\boldsymbol{L}}_S = \boldsymbol{I}_S^T \boldsymbol{L} \boldsymbol{I}_S$ consists of the rows and columns of $\boldsymbol{L}$ indexed by S and zero rows and zero columns by S^C: specifically,

$$\left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)_{ij} = \begin{cases} 0, & i,j \in S \\ (\boldsymbol{L})_{ij}, & i,j \in S^c \end{cases} \tag{9}$$

The variation difference can be further simplified as follows:

$$\begin{aligned} \left|\boldsymbol{f}^T\left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)\boldsymbol{f}\right| &= \left|\sum_{i,j} f_i f_j \left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)_{ij}\right| \\ &\le f_{\max}^2 \sum_{i,j} \left|\left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)_{ij}\right| \end{aligned} \tag{10}$$

where $f_{\max} = \max |f_i|$.

Thus, in this work, we focus on minimizing $\sum_{i,j} \left|\left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)_{ij}\right|$ alternatively, the Frobenius matrix norm $\| \boldsymbol{L} - \overline{\boldsymbol{L}}_S \|_F^2 \equiv \sum_{i,j} \left|\left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)_{ij}\right|^2$:

$$S^* = \arg\min_S \| \boldsymbol{L} - \overline{\boldsymbol{L}}_S \|_F^2 \tag{11}$$

To minimize the Frobenius matrix norm, we take a simple and fast strategy and summarize the proposed fast method in what follows:

Denoting the columns and rows of $\boldsymbol{L}$ by $\boldsymbol{c}_i$ and $\boldsymbol{r}_i$, respectively, we initially let $\boldsymbol{\Delta}_S \equiv \left(\boldsymbol{L} - \overline{\boldsymbol{L}}_S\right)$ with $\overline{\boldsymbol{L}}_S = \boldsymbol{0}$ and $S = \varnothing$.

Step 1: find the maximum norm of each column in $\boldsymbol{\Delta}_S$:

$$i^* = \arg\max_{i \in S^C} \|\boldsymbol{c}_i\|^2 \tag{12}$$

Step 2: update $\boldsymbol{\Delta}_S$ by replacing $\boldsymbol{c}_{i^*}$ and $\boldsymbol{r}_{i^*}$ with zero vectors, respectively, and let $S = S + \{i^*\}$.

Step 3: repeat Step 1 to Step 2 until S reaches the desired cardinality.

Ⅳ. Discussion of Proposed principle

In this section, we discuss that the principle of the proposed fast method is similar to that of the non-uniform sampling in [8]. First, the localization operator matrix is $\boldsymbol{T}_{\boldsymbol{g}}$ given by

$$\boldsymbol{T}_{\boldsymbol{g}} = [\, \boldsymbol{T}_1 \boldsymbol{g} \cdots \; \boldsymbol{T}_N \boldsymbol{g}] \tag{13}$$

where $\boldsymbol{T}_i \boldsymbol{g}$ is the translated version of the signal $\boldsymbol{g}$ at node i:

$$\boldsymbol{T}_i \boldsymbol{g} = \sqrt{N} \sum_{l=1}^{N} \tilde{g}(\lambda_l) \boldsymbol{u}_l^*(i) \boldsymbol{u}_l \tag{14}$$

Thus, we have

$$\boldsymbol{T}_{\boldsymbol{g}} = \sqrt{N}\, \boldsymbol{U} \widetilde{\boldsymbol{G}} \boldsymbol{U}^T, \;\; \widetilde{\boldsymbol{G}} = diag\left(\tilde{g}(\lambda_1), \ldots, \tilde{g}(\lambda_N)\right) \tag{15}$$

The non-uniform sampling strategy select nodes with the probability proportional to $\| \boldsymbol{T}_i \boldsymbol{g} \|^2$ which is the i-th column of $\boldsymbol{T}_{\boldsymbol{g}}$. To find the connection to the proposed method, we can regard $\boldsymbol{T}_{\boldsymbol{g}}$ as $\boldsymbol{L}$ with the eigenvalue matrix $\boldsymbol{\Lambda} = \sqrt{N} \widetilde{\boldsymbol{G}}$. Hence, sampling nodes corresponding to the columns with maximum norm would be similar to the non-uniform sampling method.

However, in contrast, our method differentiates itself from the non-uniform sampling by replacing the selected columns and rows with zero vectors, which is needed since the previously selected nodes should make an effect on sampling of next nodes. Furthermore, the proposed method directly operates on the variation operator $\boldsymbol{L}$ without computation of eigenvectors and eigenvalues of the operator, leading to a fast and low-weight process.

Ⅴ. Simulation results

In the experiments, we consider three different graphs for test of the different selection methods:

1) Random sensor graph (RSG)

2) Random regular graph (RRG) with each vertex connected to six vertices

3) Random Erdös-Rényi graph (RERG) with the edge connecting probability p=0.05

We generate 50 graph realizations with N=1000 vertices for each of three graphs, and create the combinatorial Laplacian matrix $\boldsymbol{L}$ as a variation operator. We construct sampling sets S with size $|S|$ from 30 to 100 by using three different techniques, denoted by efficient sampling method (ESM) [4], the proposed method (Proposed 1) given by (11) and the proposed fast method (Proposed 2), respectively. In testing the selection methods, we use random graph signals to be assumed to be generated from the Gaussian joint distribution as follows:

$$p(\boldsymbol{f}) \propto \exp(-\boldsymbol{f}^T \boldsymbol{K}^{-1} \boldsymbol{f}) = \exp(-\boldsymbol{f}^T (\boldsymbol{L}+\delta \boldsymbol{I}) \boldsymbol{f}) \quad (16)$$

where $\boldsymbol{L}+\delta \boldsymbol{I}$ is the inverse of the covariance matrix $\boldsymbol{K}$ and δ is set to be a small value (=0.01) to ensure the existence of the inverse. We generate noisy graph signals by using an iid additive noise drawn from $N(0,\sigma^2)$ and evaluate the selection methods in terms of the average reconstruction error given by $E\frac{\| \boldsymbol{f}-\hat{\boldsymbol{f}} \|^2}{N}$ in which the reconstructed signal is obtained by (3) and the average taken over 100 graph signal values at each node. In the experiments, we create graphs including $\boldsymbol{L}$ and $\boldsymbol{U}$ with an aid of the graph signal processing toolbox (GSPBox) for Matlab [11].

We investigate how the selection methods work with respect to sample size from 30 to 100 for noisy graph signals. We first provide the ratio of the running time for ESM and Proposed 2 in Table 1. Obviousy, the proposed fast algorithm runs over 10 times faster than ESM for the graphs tested. In Figure 1, 2 and 3, it is shown that the proposed fast algorithm (Proposed 2) achieves a reasonable reconstruction performance with a significantly reduced complexity as compared with ESM. More specifically, Proposed 2 shows the average reconstruction performance about 7% worse than ESM for Random sensor graph, offering a practical solution in real-time applications.

Table 1. The runninng time in second provided for ESM and Proposed 2: RT(ESM) and RT(Prop 2) represent the running time of ESM and Proposed 2, respectively.

	RSG	RRG	RERG
$\frac{RT(ESM)}{RT(\mathrm{Prop}2)}$	12.11	12.81	10.70

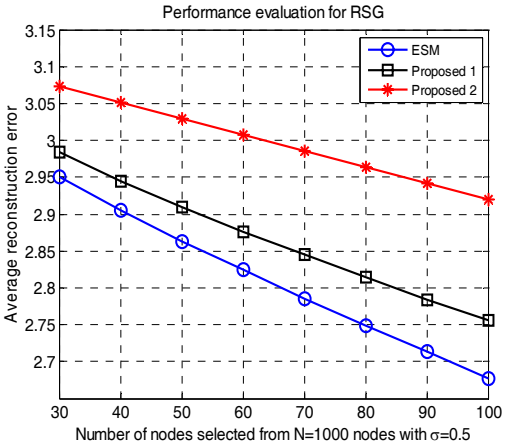

Fig. 1. Performance evaluation of different sampling methods for Random sensor graph by varying sample size with signal noise level $\sigma = 0.5$.

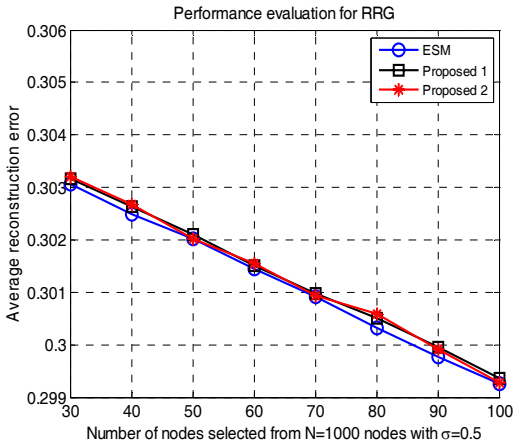

Fig. 2. Performance evaluation of different sampling methods for Random regular graphs by varying sample size with signal noise level $\sigma = 0.5$.

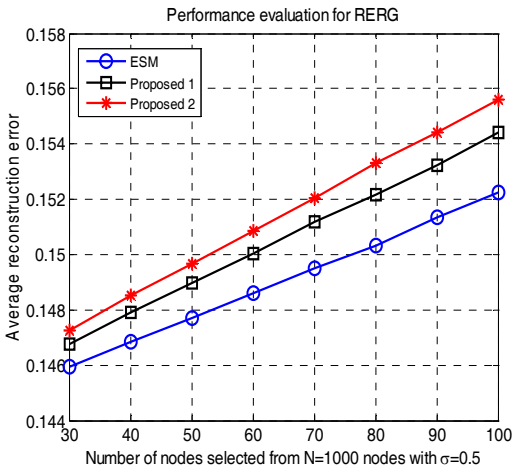

Fig. 3. Performance evaluation of different sampling methods for Random Erdös-Rényi graph by varying sample size with signal noise level $\sigma = 0.5$.

VI. Conclusion

We studied the problem of sampling a subset of nodes of graphs for arbitrary graph signals. To avoid a huge computational cost due to the eigen-decomposition of variation operators, we suggested the variation difference as a simple and indirect metric computed directly from the variation operator. We presented a greedy selection that minimizes the metric by eliminating a row and a column with the maximum norm iteratively, which corresponds to the selected node at iterations. We also discussed that the proposed algorithm operates based on the principle similar to that of the previous method [6]. We finally executed experiments to assert that the proposed algorithm offers a fast selection process with reasonable performance for various graphs as compared with the different sampling methods.

Acknowledgement

This work was supported by research fund from Chosun University (2020).

References

[1] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. "The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Mag.*, vol. 30, no. 3, 2013, pp. 83-98.

[2] A. Ortega, P. Frossard, J. Kovaevic, J. M. F. Moura, and P. Vandergheynst. "Graph signal processing: overview, challenges and applications," *Proc. of the IEEE*, vol. 106, no. 5, 2018, pp. 808-828.

[3] A. Anis, A. Gadde, and A. Ortega. "Towards a sampling theorem for signals on arbitrary graphs," *IEEE International Conf. on Acoustic, Speech, and Signal Processing (ICASSP)*, Florence, Italy, 2014, pp. 3864-3858.

[4] A. Anis, A. Gadde, and A. Ortega. "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. on Signal Processing*, vol. 64, no. 14, 2016, pp. 3775-3789.

[5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovaevic. "Discrete signal processing on graphs: sampling theory," *IEEE Trans. on Signal Processing*, vol. 63, no. 24, 2015, pp. 6510-6523.

[6] S. P. Chepuri and G. Leus, "Graph Sampling for covariance estimation," *IEEE Trans. on Signal and Information Processing over Networks*, vol. 3, no. 3, 2017, pp. 451-466.

[7] L.F.O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. on Signal Processing*, vol. 66, no. 1, 2018, pp. 34-47.

[8] N. Perraudin, B. Ricaud, D. Shuman, and P. Vandergheynst. "Global and local uncertainty principles for signals on graphs," *APSIPA Trans. On Signal and Information Processing*, vol. 7, e3, Apr., 2018, pp. 1-26.

[9] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Eigendecompostionfree sampling set selection for graph signals," *IEEE Trans. on Signal Processing*, vol. 67, no. 10, 2019, pp. 2679 - 2692.

[10] F. Wang, G. Cheung, and Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via Neumann series," *IEEE Trans. On Signal Processing*, vol. 67, no. 21, 2019, pp. 5511-5526.

[11] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond. "GSPBOX: A toolbox for signal processing on graphs," *ArXiv e-prints arXiv:1408.5781*, Aug. 2014.

저자 소개

Yoon-Hak Kim

1992: BS degree in Electronic Engineering, Yonsei University

1994: MS degree in Electronic Engineering, Yonsei University

2007: Ph.D. degree in Electrical Engineering, University of Southern California.

2012 - Present: Associate professor, Chosun Univeristy