특집논문 (Special Paper)
방송공학회논문지 제31권 제3호, 2026년 5월 (JBE Vol.31, No.3, May 2026)
https://doi.org/10.5909/JBE.2026.31.3.407
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

시계열 결측치 복원을 위한 반복적 이웃 평활화 기반 보간법

김파란하늘[a)*], 이 연 서[a)*], 현 장 훈[a)‡]

An Iterative Neighbor-Smoothing-Based Interpolation Method for Time-Series Missing Value Imputation

Paranhaneul Kim[a)*], Yeonseo Lee[a)*], and Janghun Hyeon[a)‡]

요 약

시계열 데이터의 결측치 처리는 데이터 분석의 신뢰성 확보를 위한 필수 전처리 단계이다. 그러나 기존 보간 기법들은 데이터의 특성에 따라 복원 성능 편차가 크고, 최적 기법을 사전에 판단하기 어렵다는 한계가 있다. 본 연구에서는 이러한 기법 선택 의존성을 완화하고자 기존 보간 기법에 독립적으로 적용 가능한 반복적 이웃 평활화 보간 알고리즘인 INSI를 제안한다. 제안 알고리즘은 임의의 보간 기법으로 생성된 초기 보간값에 대해 결측 시점 주변 이웃 값들의 거리 기반 가중 평균을 반복 적용하여 최종 보간값을 산출하며, 초기값에 관계없이 유일한 값으로 수렴함을 수학적으로 증명하였다. 성능 검증을 위해 NAB 데이터셋에서 선형 보간, KNN, ARIMA, BRITS, SAITS를 기저 보간으로 적용하여 단일 기법과 비교하였으며, 실험 결과, 제안 알고리즘은 MAE와 RMSE가 평균 약 68% 향상되어 결측 시점에 적용된 초기값과 무관하게 안정적인 값으로 수렴함을 확인하였다.

Abstract

Missing value imputation in time-series data is an essential preprocessing step for ensuring the reliability of data analysis. However, existing interpolation methods exhibit significant variations in restoration performance depending on data characteristics, and it remains difficult to determine the optimal method in advance. To alleviate this method selection dependency, this paper proposes INSI (Interpolation-independent Neighbor Smoothing Imputation), a post-processing interpolation algorithm based on iterative neighbor smoothing that can be applied independently of the underlying interpolation method. The proposed algorithm iteratively applies distance-based weighted averaging of neighboring values around missing time points to the initial imputed values generated by an arbitrary interpolation method, thereby producing the final imputed values. Furthermore, it is mathematically proven that the algorithm converges to a unique value regardless of the initial imputed values. For performance evaluation, experiments were conducted on the NAB dataset using linear interpolation, KNN, ARIMA, BRITS, and SAITS as base interpolation methods, and the restoration performance was compared with that of each standalone method. Experimental results demonstrate that the proposed algorithm improves MAE and RMSE by approximately 68% on average compared to the standalone methods, and converges to stable values irrespective of the base interpolation method employed.

Keyword : Time Series, Missing Value, Iterative Neighbor Smoothing, Imputation, Machine Learning, Deep Learning

Ⅰ. 서 론

제조·설비 모니터링, 에너지, 교통, 금융 등 다양한 산업 환경에서 생성되는 데이터 형식은 시간의 흐름에 따라 수집된 데이터인 시계열 데이터(Time-Series Data)이다[1]. 이러한 시계열 데이터는 실제 산업 현장에서 센서 오류, 통신 장애, 장비의 일시적 중단 등으로 인해 결측치(Missing Value)가 빈번하게 발생한다. 이렇게 생성된 결측치는 데이터의 시간적 연속성과 패턴을 훼손하며, 결측 비율이 높아질수록 데이터의 통계적 특성이 왜곡되어 데이터 분석의 정확도 및 품질을 저하시키는 요인이 된다[2]. 따라서 시계열 데이터의 결측치를 처리하는 보간(Interpolation) 기법은 데이터 분석의 신뢰성을 확보하기 위한 필수적인 전처리 단계로 간주된다.

기존 시계열 결측치 보간 방법들은 크게 통계적, 회귀 분석, 머신러닝, 딥러닝 기반 기법 등으로 구분할 수 있다. 이러한 방법들은 각각의 장점을 가지지만, 데이터의 특성과 결측 양상에 따라 복원 성능의 편차가 크게 나타난다. 이는 기존의 통계적 보간 기법부터 최신 딥러닝 기반 보간 기법에 이르기까지 어떤 보간 기법이 가장 적합한지 사전에 판단하기 어렵고, 하나의 기법이 모든 상황에서 일관되게 우수한 성능을 보장하기 어렵다는 한계가 존재한다.

본 논문에서는 이러한 보간 기법 선택 의존성 문제를 완화하기 위해 INSI(Interpolation Independent Neighbor Smoothing for Time-Series Imputation)를 제안한다. 제안 알고리즘은 기존 보간 기법을 통해 초기 보간값을 설정한 뒤, 결측 시점 주변의 이웃 값들에 대해 거리 기반 가중 평균을 반복적으로 적용하여 최종 보간값을 생성한다. 이 과정에서 반복(Iteration)이 진행됨에 따라 연속된 업데이트 값의 변화량이 점차 감소하며, 최종 보간값이 기저 보간 기법의 종류에 상관없이 하나의 안정적인 값으로 수렴하는 특징을 가진다. 제안 알고리즘의 성능을 검증하기 위해 선형 보간, KNN, ARIMA, BRITS, SAITS를 기저 보간으로 활용하여 단일 보간 기법과 제안 기법의 복원 성능을 비교하였으며, 실험 결과 모든 결측 비율에서 제안 기법이 단일 보간 기법 대비 복원 오차가 감소하였다. 또한, 기저 보간 기법 종류에 상관없이 유사한 수준의 안정적인 복원값으로 수렴함을 확인하여 제안 기법의 우수성을 증명하였다.

a) 국립한밭대학교 인공지능소프트웨어학과(Hanbat National University)
*) Equal contribution
‡ Corresponding Author : 현장훈(Janghun Hyeon)
E-mail: janghun0414@gmail.com
Tel: +82-44-863-9265
ORCID: https://orcid.org/0000-0001-5737-9179
※ 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2025년 동계학술대회에서 발표한 바 있음.
※ 본 연구는 2025년도 산업통상 자원부 및 산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임 (과제번호: RS-2024-00432506)
· Manuscript March 12, 2026; Revised April 27, 2026; Accepted April 27, 2026.

Ⅱ. 관련 연구

1. 통계적 보간 기법

1.1 전·후진 보간법(Forward·Backward Interpolation)

전진 보간법(Forward Interpolation)[2]은 결측 발생 이전 시점에서 시간적으로 가장 가까운 관측값을 그대로 대입하는 방식이며, 후진 보간법(Backward Interpolation)[2]은 결측 발생 이후 시점에서 시간적으로 가장 가까운 관측값을 사용하는 방식이다. 두 방법 모두 계산 비용이 낮고 직관적이라는 장점이 있으나, 결측 구간이 장기화될 경우 동일한 보간값을 지속적으로 유지하여 원본 데이터의 변동성을 반영하지 못한다는 한계가 존재한다.

1.2 선형 보간(Linear Interpolation)

선형 보간[3]은 결측 시점 기준 앞, 뒤 값을 직선으로 연결하여 결측값을 추정하는 방식이다. 결측 시점의 양쪽 관측값을 활용하기 때문에, 전·후진 보간법 대비 시간적 추세를 일부 반영할 수 있으며, 구현이 단순하고 계산 비용이 낮다는 장점이 있다. 그러나 두 관측값 사이의 관계를 1차 함수로 가정하므로, 데이터가 비선형적 패턴을 가질 경우 복원 정확도가 저하되는 한계가 존재한다.

1.3 스플라인 보간(Spline Interpolation)

스플라인 보간[4]은 앞서 서술한 선형 보간이 두 관측값 사이를 직선으로 연결함으로써 비선형적인 패턴을 반영하지 못하는 한계를 보완하기 위해 고안된 기법이다. 이 기법

은 보간 대상이 되는 구간을 여러 개의 구간으로 나누고 각 구간마다 2차 이상의 저차 다항식을 사용하여 곡선 형태로 결측치를 추정하는 방식이다. 가장 널리 사용되는 3차 스플라인(Cubic Spline)의 경우, 각 구간에서의 보간 함수를 3차 다항식으로 구성하며, 이를 통해 선형 보간으로는 포착하기 어려운 데이터의 비선형적인 추세 변화를 효과적으로 반영할 수 있다. 그러나 관측값이 희소하거나 결측 구간이 장기화될 경우, 다항식의 차수 특성에 의해 보간 곡선이 진동(Oscillation)하여 복원값이 실제 데이터의 범위를 벗어나는 문제가 발생할 수 있다.

1.4 K-최근접 이웃 보간(K-Nearest Neighbors Interpolation, KNN)

KNN 보간[6]은 결측이 발생한 시점으로부터 시간적으로 가장 인접한 k개의 관측 시점을 탐색하고, 해당 관측값들을 이용하여 결측값을 추정하는 기법이다. 이 기법은 크게 균등 가중(Uniform Weighted) 방식과 거리 가중(Distance Weighted) 방식으로 구분된다. 균등 가중 방식은 탐색된 이웃 관측값들에게 동일한 가중치를 부여하여 산술 평균으로 결측치를 추정하며, 거리 가중 방식은 결측 시점과 각 이웃 관측값의 시간적 거리에 반비례하는 가중치를 부여하여 시간적으로 근접한 관측값일수록 추정값에 더 큰 영향을 미치도록 설계된다. KNN 보간은 결측 시점 주변의 값들을 이용하여 결측치를 추정해 결측 시점 주변의 시간적 근접성을 반영할 수 있다는 장점이 있다. 그러나 이웃 값의 개수인 k에 따라 성능이 민감하게 변화할 수 있으며, 이웃 관측값들의 평균값을 사용하기 때문에 시계열 데이터의 변동 패턴이 올바르게 복원되지 못한다는 한계점이 존재한다.

2. 회귀 분석 기반 보간 기법

2.1 ARIMA(Autoregressive Integrated Moving Average)

회귀 분석 기법은 과거의 데이터를 이용하여 모델을 학습하고 이를 통해 결측값을 예측하는 방법이다. 대표적으로 사용되는 ARIMA 모델은 자기회귀(AR), 차분(I), 이동평균(MA)의 세 가지 요소를 결합하여 과거 데이터와 추세를 반영하여 결측치를 추정하는 모델이다. 이는 데이터의 자기 상관 구조를 반영할 수 있으며, 통계적으로 해석이 용이하다는 장점이 존재한다[5]. 그러나 시계열 데이터의 정상성(Stationarity)을 가정하고 모델링하기 때문에, 데이터에 비선형적 패턴이 많이 포함될 경우 결측치의 복원 성능이 저하되는 한계가 있다. 또한, 모델의 차수(p,d,q)를 결정하기 위해 사전에 데이터를 분석하거나 실험적으로 하이퍼파라미터를 설정해야 하는 문제가 존재한다.

3. 딥러닝 기반 보간 기법

3.1 BRITS(Bidirectional Recurrent Imputation for Time Series)

BRITS[7]는 RNN 기반의 시계열 보간 모델로, 양방향 순환 신경망 구조를 활용하여 결측 시점 이전과 이후의 관측값을 동시에 반영한다. 기존에 제안되었던 GRU-D 등의 단방향 RNN 기반 보간 기법들은 결측 시점 이전의 과거 정보만을 사용하였으나, BRITS는 순방향과 역방향의 정보를 결합함으로써 보간 정확도를 향상시켰다. 그러나, RNN의 구조 특성상 시계열 데이터의 길이가 길어질수록 장기 의존성을 포착하는데 한계가 존재하며, 모델 학습 과정에서 기울기 소실(Gradient Vanishing) 문제가 발생할 수 있다는 한계가 있다[10].

3.2 SAITS(Self-Attention-based Imputation for Time Series)

최근 자기주의 메커니즘이 발전하면서 시계열 데이터의 결측치 처리에도 사용되는 사례가 많아졌다. 대표적으로 SAITS[8]는 트랜스포머 기반의 시계열 보간 모델로 시계열 데이터의 관측값 간의 전역적 상관관계를 포착하며, 두 개의 DMSA(Diagonally Masked Self-Attention)를 결합한 가중 학습을 통해 결측치 보간과 관측값 복원을 동시에 수행한다. 이는 RNN 기반의 모델과 달리 시계열 데이터를 순차적으로 처리할 필요 없이 병렬 연산이 가능하며, RNN의 한계점인 장기 의존성 문제를 효과적으로 해결할 수 있다는 장점이 있다[11]. 그러나, 트랜스포머 구조의 특성상 모델의 파라미터 수가 많아 높은 계산 비용이 요구되며, 소규모 데이터셋에서는 과적합이 발생할 수 있다는 한계가 존재한다.

3.3 GAIN(Generative Adversarial Imputation Nets)

GAIN[12]은 GAN(Generative Adversarial Network)의 생성자(Generator)가 데이터를 생성하고 판별자(Discriminator)가 데이터를 판별하는 적대적 생성 기법을 결측치 보간에 적용한 모델로, 생성자가 결측값을 생성하고 판별자가 해당 값이 실제 관측값인지, 생성된 값인지를 판별하는 적대적 학습을 통해 보간을 수행한다. 이때 판별자에게 결측치에 대한 부분을 벡터 형식으로 제공하고, 이 정보를 통해 생성자는 실제 데이터 분포를 따라 데이터를 생성하도록 학습한다. 이러한 방식은 데이터의 분포적 특성을 더 잘 보존할 수 있다는 장점이 있으나, GAN 구조의 고질적인 문제인 학습 불안정성과 모드 붕괴(Mode Collapse) 현상이 발생할 수 있으며, 시계열 데이터의 시간적 순서 정보를 반영하지 못한다는 한계가 존재한다.

3.4 CSDI(Conditional Score-based Diffusion Models for Imputation)

CSDI[13]는 디퓨전(Diffusion) 모델을 시계열 결측치 보간에 적용한 기법으로, 관측값을 조건으로 하여 결측 시점의 데이터를 점진적으로 복원하는 조건부 생성 모델이다. 디퓨전 모델은 원본 데이터에 단계적으로 노이즈를 추가하는 순방향 과정(Forward Process)과 노이즈로부터 원본 데이터를 복원하는 역방향 과정(Reverse Process)으로 구성되며, 역방향 과정에서 관측값을 조건 정보로 활용하여 결측치의 분포를 추정한다. 즉, 기존 보간 기법들이 하나의 보간값을 출력하는 것과 달리 확률적으로 다수의 보간값을 샘플링할 수 있어 보간의 불확실성을 정량적으로 추정할 수 있으며 이를 통해 기존의 GAN 기반 방식보다 더 안정적이고 현실적인 데이터를 생성할 수 있다는 장점이 있다. 그러나 역방향 과정에서 다수의 노이즈 제거 단계를 반복해야 하므로 추론 속도가 느리며, 일반적으로 디퓨전 기반 모델은 데이터의 양이 적을 경우 모델이 과적합될 수 있다는 문제점이 존재한다.

Ⅲ. 제안 알고리즘

1. 알고리즘 설계

제안 알고리즘은 기존 보간 기법으로 생성된 초기 보간값에 대해 이웃 거리 기반 가중 평균을 반복적으로 적용하여 최종 보간값을 생성한다. 그림 1은 제안 알고리즘의 전체 흐름을 나타내며, 알고리즘 1에 동작 절차를 수도코드로 정리하였다. 결측 시점 t에 대해 좌우 k개 이웃 시점을 포함하는 윈도우 $W_t = \{t-k,\cdots,t-1,t,t+1,\cdots,t+k\}$를 정의할 때, 이웃 시점 $t-k$부터 $t+k$까지에 대한 가중치 w_h는 식 (1)과 같이 정의된다.

$$w_h = \frac{1}{h}\,(h = 1,2,\cdots,k) \tag{1}$$

초기 보간값 $\hat{X}$를 바탕으로, 각 결측 시점 t에 대한 업데이트 값 $\hat{X}_t$는 식 (2)와 같이 좌우 이웃 값의 가중 평균으로 계산된다.

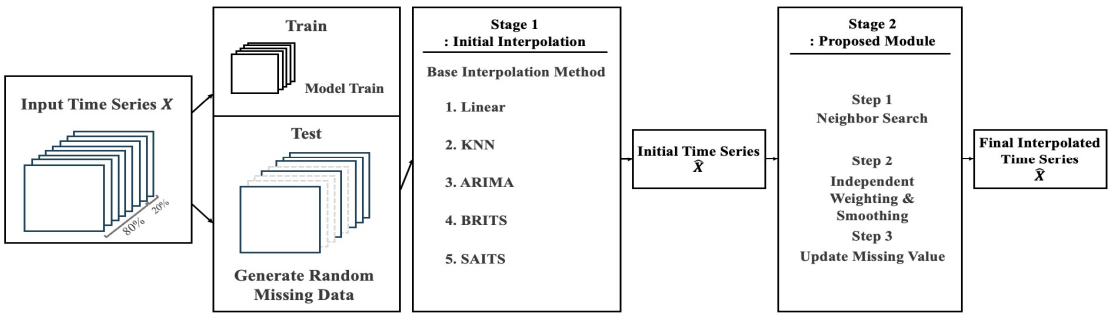


그림 1. 제안 알고리즘(INSI)의 프레임워크
Fig. 1. Overall Framework of the Proposed Algorithm(INSI)

알고리즘 1. 반복적 이웃 평활화 기반 시계열 보간 알고리즘
Algorithm 1. Proposed Iterative Neighbor Smoothing Algorithm for Time-Series Imputation

Algorithm 1 Proposed Iterative Neighbor Smoothing Algorithm for Time-Series Imputation

$\hat{X} \leftarrow \text{BASEINTERPOLATION}(X)$
Define the set of missing indices M
Set window size $2k$, convergence threshold ϵ, and maximum iteration T
$iter \leftarrow 0$
repeat
  $\hat{X}_{\text{new}} \leftarrow \hat{X}$
  for all $t \in M$ **do**
    $W_t \leftarrow \{t-k, \ldots, t-1, t+1, \ldots, t+k\}$
    for $h = 1$ to k **do**
      $d_h \leftarrow h$
      $w_h \leftarrow \frac{1}{d_h}$
    end for
    $\hat{X}_{\text{new},t} \leftarrow \frac{\sum_{h=1}^{k} w_h \hat{X}_{t-h} + \sum_{h=1}^{k} w_h \hat{X}_{t+h}}{2\sum_{h=1}^{k} w_h}$
  end for
  $\hat{X} \leftarrow \hat{X}_{\text{new}}$
  $iter \leftarrow iter + 1$
until $\|\hat{X} - \hat{X}_{\text{new}}\| < \epsilon$ **or** $iter \geq T$
return $\hat{X}$

$$\hat{X}_t = \frac{\sum_{h=1}^{k} w_h \hat{X}_{t-h} + \sum_{h=1}^{k} w_h \hat{X}_{t+h}}{2\sum_{h=1}^{k} w_h} \tag{2}$$

윈도우 내에 결측 시점이 여러 개일 경우, 업데이트 과정에서 서로의 값이 영향을 주므로 반복 과정을 통해 하나의 안정적인 값으로 수렴한다.

2. 이론적 분석

본 절에서는 제안 알고리즘이 임의의 초기 보간값에 관계없이 하나의 유일한 값으로 수렴함을 수학적으로 증명한다. 이는 기저 보간 기법의 선택이 최종 복원 성능에 영향을 미치지 않음을 의미하며, 제안 알고리즘이 기존 보간 기법에 독립적인 알고리즘이라는 이론적 근거를 제시한다.

2.1 표기법 및 업데이트 식의 일반화

윈도우 크기 $2k$에 포함되는 임의의 결측 시점이 m개 존재할 때 결측 시점의 집합을 $M = \{t_1, t_2, \ldots, t_m\}$로 정의한다. n번째 반복에서 결측 시점 t_i의 값을 $x_{t_i}^{(n)}$이라고 하고, 식 (1)과 (2)에서 정의한 업데이트 식에 의해 거리를 h, 가중치를 w_h로 정의한다. 계산의 편의성을 위해 업데이트 식의 분모를 식 (3)으로 정의하고 전체 업데이트 식을 식 (4)로 재정의한다.

$$S = 2\sum_{h=1}^{k} w_h \tag{3}$$

$$\hat{X}_t^{(n+1)} = \frac{\sum_{h=1}^{k} w_h \hat{X}_{t-h}^{(n)} + \sum_{h=1}^{k} w_h \hat{X}_{t+h}^{(n)}}{S} \tag{4}$$

또한, 결측 시점 t_i의 결측값 $x_{t_i}^{(n)}$을 업데이트할 때, 윈도우 안의 $2k$개의 이웃들 중 일부는 관측값(원본 데이터)이고 일부는 결측치를 처리한 보간값이다. 따라서 이를 식 (5)로 다시 정의할 수 있다.

$$\hat{X}_t^{(n+1)} = \frac{\text{관측값들의 가중합} + \text{보간값들의 가중합}}{S} \tag{5}$$

결측 시점 t_i의 결측값 $x_{t_i}^{(n)}$을 업데이트할 때 윈도우 내부에 또 다른 결측 시점 t_j가 존재하면, 식 (5)의 가중치 w_h에 대하여 식 (6)으로 정의할 수 있다.

$$u_{ij} = \begin{cases} w_{|t_i - t_j|}, others \\ 0, \quad i = j \end{cases} \tag{6}$$

$i = j$일 때 u_{ij}의 값이 0인 이유는, 결측 시점 t_i의 업데이트 식인 식 (6)에서 자기 자신 t_i의 값은 윈도우에 포함되지 않기 때문이다. 따라서 보간값들의 가중합을 포함하여 업데이트 식을 식 (7)로 정의할 수 있다.

계산의 편의성을 위해 iteration 횟수 n에 따라서 변하지 않는 상수항 c_i를 식 (8)로 정의한다.

$$c_i = \frac{\text{관측값들의 가중합}}{S} \tag{8}$$

다음으로 보간값 $x_{t_i}^{(n)}$ 앞에 붙는 정규화된 가중치 a_{ij}를 식 (9)로 정의한다.

$$a_{ij} = \frac{u_{ij}}{S} \tag{9}$$

따라서 식 (9)는 다음과 같이 식 (10)으로 정의할 수 있다.

$$\hat{X}_t^{(n+1)} = c_i + a_{i1}x_1^{(n)} + a_{i2}x_2^{(n)} \ldots + a_{im}x_m^{(n)} \tag{10}$$

2.2 업데이트 식의 성질

식 (9)에서 정의한 a_{ij}는 다음과 같은 성질을 가진다.

① 가중치 w_h가 양수이므로, u_{ij}도 양수이다. 또한, $S > 0$이므로 식 (11)이 성립된다.

$$a_{ij} \geq 0 \tag{11}$$

② 각 윈도우 안에 적어도 하나 이상의 관측값이 있다고 가정할 때, 식 (12)가 성립된다.

$$a_{i1} + a_{i2} + \ldots + a_{im} < 1 \tag{12}$$

모든 i에 대하여 식 (12)가 성립하므로, 식 (13), 식 (14)를 정의할 수 있다.

$$r = \max_{1 \leqq i \leqq m} \sum_{j=1}^{m} a_{ij} \tag{13}$$

$$0 \leq r < 1 \tag{14}$$

2.3 초기 보간값 독립성 증명

초기 보간값이란, 제안 알고리즘을 적용하기 이전에 기존 보간 기법을 통해 결측치를 처리한 값을 의미한다. 예를 들어, 선형 보간을 기저 보간으로 사용하면 선형 보간에 의해 추정된 보간값이 초기 보간값이 되며, SAITS를 기저 보간으로 사용하면 SAITS에 의해 추정된 값이 초기 보간값이 된다. 이때 동일한 결측 데이터에 대하여 서로 다른 두 기저 보간 알고리즘을 각각 적용하면 기저 보간 알고리즘의 복원 방식이 상이하므로 결측 시점에 대입되는 초기 보간값 또한 서로 다르게 생성된다. 본 절에서는 이렇게 생성된 서로 다른 두 초기 보간값에 대해 각각 식 (10)으로 업데이트를 수행하였을 때, iteration이 진행됨에 따라 기저 보간 알고리즘의 종류에 관계없이 동일한 값으로 수렴함을 증명한다. 이때 첫 번째 알고리즘에 의한 업데이트 결과를 $x_1^{(n)}, x_2^{(n)}, \ldots, x_m^{(n)}$, 두 번째 알고리즘에 의한 업데이트 결과를 $y_1^{(n)}, y_2^{(n)}, \ldots, y_m^{(n)}$라고 하면, 실행된 두 개의 보간값 차이는 $d_i^{(n)} = x_i^{(n)} - y_i^{(n)}$로 정의할 수 있다. 이때 $x_i^{(n)}$, $y_i^{(n)}$은 식 (10)에 의해 업데이트된 $\hat{X}_t^{(n+1)}$를 의미하므로,

$$\hat{X}_t^{(n+1)} = \frac{\text{관측값들의 가중합} + u_{i1}x_1^{(n)} + u_{i2}x_2^{(n)} + \ldots + u_{im}x_m^{(n)}}{S} \tag{7}$$

상수항을 제거하면 $d_i^{(n)}$에 대하여 식 (15)를 정의할 수 있다.

$$d_i^{(n+1)} = a_{i1}d_1^{(n)} + a_{i2}d_2^{(n)} + \ldots + a_{im}d_m^{(n)} \quad (15)$$

Iteration의 n번째 반복에서 $d_i^{(n)}$에 대하여 절댓값이 가장 큰 것을 식 (16)으로 정의한다.

$$D_n = \max(|d_1^{(n)}|, |d_2^{(n)}|, \ldots, |d_n^{(n)}|) \quad (16)$$

식 (16)에 의해 모든 n에 대하여 $|d_i^{(n)}| \leq D_n$이 성립하므로, 식 (15)를 다음과 같이 식 (17)로 정의할 수 있다.

$$|d_i^{(n+1)}| \leq a_{i1}|d_1^{(n)}| + a_{i2}|d_2^{(n)}| + \ldots + a_{im}|d_m^{(n)}| \quad (17)$$

이를 각 항에 대해서 다시 정리하면 식 (18)을 도출할 수 있다.

$$|d_i^{(n+1)}| \leq (a_{i1} + a_{i2} + \ldots + a_{im})D_n \quad (18)$$

이때 식 (12)에 의해서 식 (18)을 다시 정리하면, $|d_i^{(n+1)}| \leq rD_n$을 도출할 수 있다. 이는 모든 i에 대해 성립하므로 식 (19)를 정의할 수 있다.

$$D_{n+1} \leq rD_n \quad (19)$$

Iteration이 반복되면서 다음과 같은 식 (20)을 도출할 수 있다.

$$D_{n+p} \leq r^pD_n \quad (p \in N) \quad (20)$$

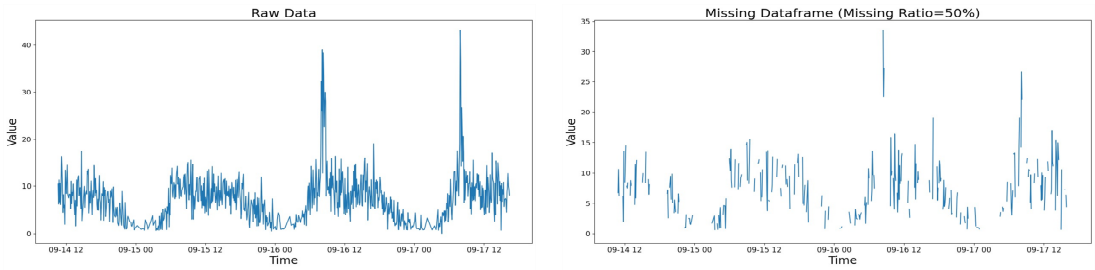
이때 식 (14)에 의해 $0 \leq r < 1$이 성립하므로 식 (21)을 도출할 수 있다.

$$0 \leq \lim_{p\to\infty} D_{n+p} \leq \lim_{p\to\infty} r^pD_{n+p-1} = 0 \quad (21)$$

즉, 모든 i에 대하여 $\lim_{n\to\infty} D_n = 0$이 성립하므로, 제안 알고리즘은 iteration이 반복되면서 초기 보간값과 무관한 유일한 값으로 수렴함을 증명할 수 있다.

Ⅳ. 실험 및 결과

본 장에서는 제안 알고리즘의 성능을 다양한 조건에서 비교·분석한다. 4.1절에서는 실험에 사용한 데이터셋, 결측 생성 방법 등 실험 환경 및 평가 지표에 대하여 설명한다. 4.2절에서는 기존 보간 기법에 INSI를 후처리 보간 방식으로 적용한 뒤, 모든 데이터셋에 대해 윈도우 크기(k)를 2, 4, 6으로 변화시키며 복원 성능을 비교한다. 4.3절에서는 결측 구간의 초기값을 임의의 랜덤한 상수로 설정한 후, INSI를 단독 보간 알고리즘으로 적용하였을 때의 복원 성능을 분석한다.



(a) Original data of occupancy_t4013 (b) occupancy_t4013 with 50% missing values

그림 2. occupancy_t4013 원본 및 결측 데이터 시각화
Fig. 2. Visualization of the Original and Missing Data in the occupancy_t4013

1. 실험 환경 및 평가 지표

본 논문에서는 제안 알고리즘의 성능을 확인하기 위해 데이터가 공개된 단변량 시계열 데이터인 NAB(Numenta Anomaly Benchmark) 데이터셋을 활용하였다. 실험에는 occupancy_t4013, Ambient temperature system failure, Speed 6005의 세 가지 데이터를 사용하였다. 이들 데이터는 각각 서로 다른 변동성과 국소 패턴을 가지므로, 제안 알고리즘의 후처리 보간 성능과 윈도우 크기 변화에 따른 특성을 함께 분석하기에 적합하다. 각 데이터셋별 원본 데이터와 결측이 발생한 데이터는 그림 2로 제시하였으며, 이를 통해 데이터셋별 결측 발생 전후의 형태 변화를 시각적으로 확인할 수 있다. 비교 기법으로는 통계적 보간 기법으로 선형 보간과 KNN을 사용하였으며, 회귀 분석 기법으로 ARIMA, 딥러닝 보간 기법으로 BRITS, SAITS를 사용하여 단일 보간 알고리즘과 제안 알고리즘의 후처리 보간 성능을 각각 비교하였다. 이때 선형 보간, KNN은 학습 과정 없이 직접 보간을 수행하는 반면, ARIMA, BRITS, SAITS는 학습 데이터를 기반으로 결측치를 추정한다. 따라서 공정한 성능 평가를 위해 데이터의 80%를 학습 데이터셋으로 사용하고 20%를 테스트 데이터셋으로 사용하였으며, 테스트 데이터셋에 대해서 10%~50% 비율로 무작위 결측을 발생시켜 복원 성능을 평가하였다. 평가 지표로는 식 (22)의 평균 절대오차(MAE)와 식 (23)의 평균 제곱근오차(RMSE)를 사용하였다. MAE는 복원값과 실제값 사이의 절대 오차 평균을 나타내며, RMSE는 복원값과 실제값 사이의 제곱 오차 평균에 제곱근을 취한 값으로, 큰 오차에 더 민감하게 반응하는 지표이다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\widehat{Y}_i - Y_i| \tag{22}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|\widehat{Y}_i - Y_i|} \tag{23}$$

2. 후처리 보간에서의 윈도우 크기별 성능 비교

본 절에서는 기존 보간 기법으로 생성된 초기 보간값에 제안 알고리즘을 후처리 보간 방식으로 적용한 뒤, 윈도우 크기(k)에 따른 복원 성능 차이를 비교한다. 이를 위해 제안 알고리즘의 k값을 각각 2, 4, 6으로 설정하여 각 데이터셋의 결측 비율별 복원 성능을 측정하였으며, 실험 결과 윈도우 크기에 따른 복원 성능은 데이터의 특성에 따라 상이하게 나타났다. occupancy_t4013 데이터셋은 일정 구간 동안 유사한 패턴을 가지므로, 결측 시점 주변의 값만 보는 것보다 넓은 구간의 상태 정보를 반영하는 것이 복원에 더 유리하게 작용한다. 이에 따라 윈도우 크기 k가 증가할수록 복원 성능이 개선되는 경향이 나타났으며, 특히 결측 비율이 증가할수록 이러한 경향은 더욱 뚜렷하게 확인되었다. 표 1에서 확인할 수 있듯이 기존 단일 보간 기법보다 제안 알고리즘이 전반적으로 우수한 성능을 보였으며, 이는 일정 구간 동안 유사한 패턴이 유지되는 데이터에서는 제안 알고리즘이 기존 보간 기법보다 더 효과적으로 결측 구간

표 1. occupancy_t4013 데이터셋의 무작위 결측 비율 및 윈도우 크기에 따른 보간 성능 비교
Table 1. Comparison of interpolation performance according to random missing rates and window sizes on the occupancy_t4013 dataset

	Missing Rate(%)	Linear	Linear+INSI			KNN	KNN+INSI			ARIMA	ARIMA+INSI			BRITS	BRITS+ISNI			SAITS	SAITS+INSI		
	Window Size	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6
MAE	10	2.71	2.51	2.54	**2.50**	**2.50**	2.51	2.54	**2.50**	2.74	2.51	2.54	**2.50**	5.68	2.51	2.54	**2.50**	3.97	2.51	2.54	**2.50**
	30	2.79	2.56	2.45	**2.42**	2.47	2.56	2.45	**2.42**	2.63	2.56	2.45	**2.42**	5.15	2.56	2.45	**2.42**	4.85	2.56	2.45	**2.42**
	50	2.96	2.82	2.70	**2.64**	2.84	2.82	2.70	**2.64**	3.01	2.82	2.70	**2.64**	5.12	2.82	2.70	**2.64**	4.10	2.82	2.70	**2.64**
RMSE	10	3.40	**3.19**	3.39	3.39	3.25	**3.19**	3.39	3.39	3.72	**3.19**	3.39	3.39	7.78	**3.19**	3.39	3.39	6.01	**3.19**	3.39	3.39
	30	3.69	3.33	**3.25**	**3.25**	3.27	3.33	**3.25**	**3.25**	3.75	3.33	**3.25**	**3.25**	6.91	3.33	**3.25**	**3.25**	6.60	3.33	**3.25**	**3.25**
	50	4.32	4.11	3.90	**3.84**	4.31	4.11	3.90	**3.84**	4.72	4.11	3.90	**3.84**	7.20	4.11	3.90	**3.84**	6.03	4.11	3.90	**3.84**

표 2. Ambient temperature system failure 데이터셋의 무작위 결측 비율 및 윈도우 크기에 따른 보간 성능 비교
Table 2. Comparison of interpolation performance according to random missing rates and window sizes on the Ambient temperature system failure dataset

	Missing Rate(%)	Linear	Linear+INSI			KNN	KNN+INSI			ARIMA	ARIMA+INSI			BRITS	BRITS+ISNI			SAITS	SAITS+INSI		
	Window Size	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6
MAE	10	0.64	**0.61**	0.63	0.69	0.62	**0.61**	0.63	0.69	0.86	**0.61**	0.63	0.69	61.91	**0.61**	0.63	0.69	51.37	**0.61**	0.63	0.69
	30	0.61	**0.60**	0.65	0.74	0.71	**0.60**	0.65	0.74	0.91	**0.60**	0.65	0.74	62.62	**0.60**	0.65	0.74	49.65	**0.60**	0.65	0.74
	50	0.67	**0.66**	0.73	0.85	0.86	**0.66**	0.73	0.85	1.06	**0.66**	0.73	0.85	60.65	**0.66**	0.73	0.85	49.85	**0.66**	0.73	0.85
RMSE	10	0.77	**0.75**	0.80	0.88	0.76	**0.75**	0.80	0.88	1.12	**0.75**	0.80	0.88	62.02	**0.75**	0.80	0.88	51.48	**0.75**	0.80	0.88
	30	**0.75**	**0.75**	0.83	0.95	0.89	**0.75**	0.83	0.95	1.16	**0.75**	0.83	0.95	62.72	**0.75**	0.83	0.95	49.77	**0.75**	0.83	0.95
	50	**0.85**	**0.85**	0.95	1.11	1.10	**0.85**	0.95	1.11	1.35	**0.85**	0.95	1.11	60.76	**0.85**	0.95	1.11	49.97	**0.85**	0.95	1.11

을 복원할 수 있음을 의미한다.

Ambient temperature system failure 데이터셋에서는 occupancy_t4013과는 상이한 결과가 나타났다. 해당 데이터는 국소적인 변동이 반복적으로 나타나는 동시에 일부 구간에서 급격한 상승 및 하강을 포함하고 있어, 결측 구간 복원 시 넓은 범위의 정보를 반영하는 것보다 결측 시점 주변의 국소 패턴을 보존하는 것이 중요하다. 실험 결과, 표 2에서 확인할 수 있듯이 윈도우 크기 $k=2$에서 제안 알고리즘이 우수한 성능을 보였으나, k값이 증가할수록 복원값이 과도하게 평활화되어 성능이 저하되었다. 이는 국소적으로 급격한 변화를 포함하는 데이터에서는 윈도우 크기에 따라 복원 성능이 크게 달라질 수 있음을 시사하며, 따라서 이러한 경우에는 큰 윈도우보다 작은 크기의 윈도우를 사용하는 것이 더 적합함을 의미한다.

speed_6005 데이터셋은 값의 변화가 크고 불규칙한 변동이 반복적으로 나타나는 특성을 가지고 있어, 결측 구간 복원 시 좁은 범위의 이웃 값만 반영할 경우 원본 데이터의 변화 양상을 충분히 따라가기 어렵고, 반대로 지나치게 넓은 범위를 반영할 경우 급격한 변화 구간까지 평활화되어 복원 성능이 저하될 수 있다. 실험 결과, 표 3에서 확인할 수 있듯이 전반적으로 k가 커질수록 더 우수한 복원 성능이 나타났으며, 특히 $k=4$일 때, 모든 결측 비율에서 가장 일관되고 안정적인 결과를 보였다. 이는 불규칙한 패턴을 가진 데이터에서는 원본 데이터의 변화 양상을 반영할 수 있는 적절한 크기의 윈도우를 설정하는 것이 중요함을 의미한다.

종합하면, k는 제안 알고리즘의 성능을 좌우하는 하이퍼파라미터이다. 이는 제안 알고리즘이 결측 시점 주변 이웃 값들의 가중 평균을 반복적으로 반영하므로, k값이 작을수록 국소적인 정보에 민감하게 반응하고, k값이 커질수록 넓은 구간의 정보를 반영하기 때문이다. 따라서 제안 알고리즘이 항상 기존 보간 기법보다 우수한 성능을 보장하는 것은 아니며, 데이터의 특성에 따라 적절한 윈도우 크기 k

표 3. speed_6005 데이터셋의 무작위 결측 비율 및 윈도우 크기에 따른 보간 성능 비교
Table 3. Comparison of interpolation performance according to random missing rates and window sizes on the speed_6005 dataset

	Missing Rate(%)	Linear	Linear+INSI			KNN	KNN+INSI			ARIMA	ARIMA+INSI			BRITS	BRITS+ISNI			SAITS	SAITS+INSI		
	Window Size	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6	-	2	4	6
MAE	10	6.37	6.14	6.08	**6.01**	6.18	6.14	6.08	**6.01**	**5.85**	6.14	6.08	6.01	79.96	6.14	6.08	**6.01**	72.85	6.14	6.08	**6.01**
	30	6.56	6.31	6.29	**6.28**	6.47	6.31	6.29	**6.28**	6.91	6.31	6.29	**6.28**	79.37	6.31	6.29	**6.28**	73.55	6.31	6.29	**6.28**
	50	6.87	6.52	**6.37**	6.43	6.82	6.53	**6.37**	6.43	7.62	6.52	**6.37**	6.43	79.54	6.63	**6.38**	6.43	74.70	6.62	**6.38**	6.43
RMSE	10	8.00	7.83	7.68	**7.60**	7.94	7.83	7.68	**7.60**	**7.33**	7.83	7.68	7.60	80.46	7.83	7.68	**7.60**	73.40	7.83	7.68	**7.60**
	30	8.46	8.07	7.96	**7.91**	8.11	8.07	7.96	**7.91**	8.64	8.07	7.96	**7.91**	79.89	8.07	7.96	**7.91**	74.11	8.07	7.96	**7.91**
	50	8.87	8.39	**8.10**	8.12	8.52	8.39	**8.10**	8.12	10.67	8.39	**8.10**	8.12	80.08	8.51	**8.11**	8.12	75.26	8.49	**8.11**	8.12

를 설정하는 것이 중요하다. 한편, BRITS와 SAITS와 같은 딥러닝 기반 보간 기법은 본 연구에서 전반적으로 낮은 복원 성능을 보였는데, 이는 본 연구에서 사용한 데이터가 단변량 시계열 데이터이며, 학습 데이터의 양 또한 충분하지 않아 딥러닝 모델이 복잡한 시계열 구조를 안정적으로 학습하기 어려웠기 때문으로 분석된다.

3. 무작위 상수 초기값 기반 단일 보간 실험

본 절에서는 제안 알고리즘이 기존 보간 기법의 후처리 방식에 한정되지 않고, 결측 구간의 초기값을 무작위 상수로 설정한 경우에도 단독 보간 알고리즘으로 활용될 수 있는지를 검증한다. 이를 위해 각 데이터셋의 결측 구간 초기값을 해당 데이터셋의 최솟값과 최댓값 사이의 임의의 상수값으로 설정한 뒤, 제안 알고리즘을 적용하여 복원 성능과 실행 시간을 측정하였다. 실험 결과, 무작위 상수 초기값을 사용한 경우에도 제안 알고리즘은 모든 데이터셋에서 안정적인 복원 성능을 보였다. 표 4~5의 Ambient_temperature_system_ failure, speed_6005 데이터셋에서는 모든 결측 비율에서 가장 낮은 MAE를 기록하였으며, RMSE 역시 대부분의 구간에서 가장 낮거나 선형 보간과 동일한 복원 성능을 나타냈다. 반면, 표 6의 occupancy_t4013에서는 일부 결측 비율에서 KNN이 더 낮은 MAE를 기록하였으나, 제안 알고리즘 역시 우수한 성능을 유지하였으며, RMSE 기준으로는 대부분의 결측 비율에서 가장 우수한 복원 성능을 보였다. 이러한 결과는 제안 알고리즘이 초기값 설정에 크게 영향 받지 않으며, 서로 다른 특성을 가지는 데이터에서도 안정적인 복원 성능을 나타낼 수 있음을 의미한다. 따라서, 제안 알고리즘은 기존 보간 기법의 후처리 단계로 한정되지 않고, 초기값에 대한 사전 정보가 충분하지 않은 상황에서도 단독 보간 알고리즘으로 활용될 수 있다. 한편, 실행 시간 측면에서는 선형 보간 및 KNN보다 다소 길었으나, 표 4~6에 나타난 바와 같이 ARIMA, BRITS, SAITS에 비하여 현저히 짧은 수행 시간을 보였다. 이는 제안 알고

표 4. Ambient temperature system failure 데이터셋의 무작위 결측 비율 및 보간 소요 시간에 따른 성능 비교
Table 4. Comparison of Performance According to the Random Missing Rate and Imputation Time on the Ambient temperature system failure Dataset

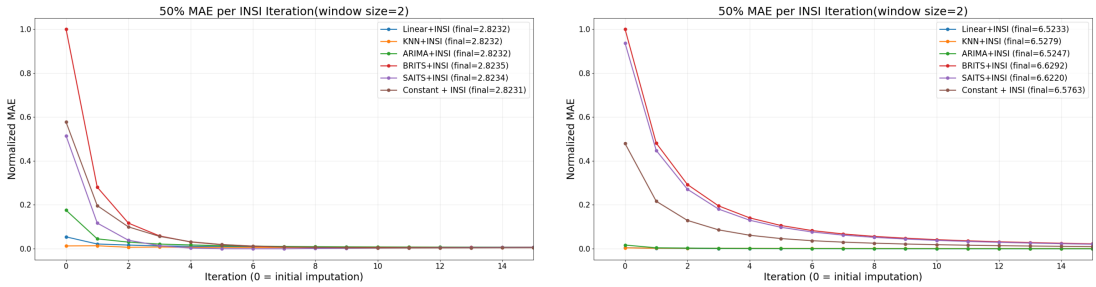
	MAE					RMSE					Time(s)				
Missing Rate (%)	10	20	30	40	50	10	20	30	40	50	10	20	30	40	50
Linear	0.63	0.61	0.60	0.61	**0.66**	0.77	0.74	0.75	0.77	**0.84**	**0.0004**	**0.0004**	**0.0005**	**0.0004**	**0.0004**
KNN	0.61	0.66	0.70	0.76	0.85	0.76	0.84	0.89	0.97	1.10	0.0016	0.0030	0.0045	0.0061	0.0078
ARIMA	0.85	0.86	0.91	0.96	1.05	1.11	1.09	1.16	1.26	1.35	0.3259	0.5270	0.9715	1.0245	0.5259
BRITS	60.37	61.27	60.03	62.70	62.50	60.49	61.37	60.15	62.80	62.60	23.3672	22.4708	21.5117	19.4754	25.0951
SAITS	57.52	57.09	56.70	54.50	52.14	57.63	57.19	56.80	54.61	52.25	3.2145	2.0340	1.9760	1.9810	1.9645
INSI	**0.60**	**0.59**	**0.59**	**0.61**	**0.66**	**0.74**	**0.74**	**0.75**	**0.77**	**0.84**	0.0289	0.0916	0.1362	0.1800	0.2269

표 5. speed_6005 데이터셋의 무작위 결측 비율 및 보간 소요 시간에 따른 성능 비교
Table 5. Comparison of Performance According to the Random Missing Rate and Imputation Time on the speed_6005 Dataset

	MAE					RMSE					Time(s)				
Missing Rate (%)	10	20	30	40	50	10	20	30	40	50	10	20	30	40	50
Linear	6.36	6.33	6.55	6.64	6.86	7.99	8.19	8.46	8.51	8.86	**0.0004**	**0.0004**	**0.0004**	**0.0004**	**0.0004**
KNN	6.18	6.14	6.47	6.60	6.81	7.93	7.75	8.11	8.26	8.52	0.0010	0.0019	0.0028	0.0038	0.0048
ARIMA	5.85	6.49	6.90	7.06	7.61	7.32	8.07	8.64	8.91	10.67	0.5876	0.5240	0.5723	0.7242	0.4236
BRITS	80.02	81.19	78.95	79.04	78.99	80.53	81.63	79.48	79.57	79.53	12.1563	9.3717	9.2223	10.4026	9.8940
SAITS	78.29	76.37	76.11	77.06	75.89	78.81	76.84	76.65	77.60	76.45	1.5264	0.9466	1.0059	0.9419	0.9704
INSI	**6.13**	**6.03**	**6.30**	**6.42**	**6.57**	**7.82**	**7.73**	**8.07**	**8.17**	**8.43**	0.0191	0.0401	0.0819	0.1095	0.1349

표 6. occupancy_t4013 데이터셋의 무작위 결측 비율 및 보간 소요 시간에 따른 성능 비교
Table 6. Comparison of Performance According to the Random Missing Rate and Imputation Time on the occupancy_t4013 Dataset

	MAE					RMSE					Time(s)				
Missing Rate (%)	10	20	30	40	50	10	20	30	40	50	10	20	30	40	50
Linear	2.70	2.66	2.79	2.74	2.95	3.39	3.56	3.69	3.85	4.31	**0.0004**	**0.0004**	**0.0004**	**0.0004**	**0.0004**
KNN	**2.50**	**2.45**	**2.46**	**2.53**	2.84	3.24	3.29	**3.27**	3.53	4.30	0.0008	0.0018	0.0026	0.0035	0.0045
ARIMA	2.74	2.67	2.62	2.72	3.01	3.71	3.78	3.75	4.03	4.72	0.5851	0.4914	0.6482	0.5565	0.4438
BRITS	5.05	5.35	5.29	5.70	5.93	7.17	7.43	7.07	7.44	7.93	11.6001	9.2903	10.1209	10.4881	9.4693
SAITS	4.99	4.38	4.01	4.44	4.75	7.08	6.41	5.64	6.10	6.75	0.9810	0.9794	1.1479	1.3850	1.2927
INSI	**2.50**	2.51	2.55	2.56	**2.82**	**3.19**	**3.29**	3.32	**3.49**	**4.10**	0.0170	0.0524	0.0767	0.1369	0.1262



(a) occupancy_t4013 (k=2) (b) speed_6005 (k=2)

그림 3. 데이터별 iteration에 따른 복원 오차 변화
Fig. 3. Changes in restoration error with iteration by dataset

리즘이 복잡한 학습 과정 없이도 안정적인 복원 성능을 확보할 수 있음을 보여주며, 데이터의 규모가 제한적이거나 계산 효율이 중요한 실제 환경에서 유용하게 활용될 수 있음을 시사한다.

그림 3은 각 데이터셋에서 다양한 윈도우 크기 조건에서 iteration이 진행됨에 따라 MAE가 감소하는 양상을 나타낸다. 초기 단계에서는 기저 보간 기법의 종류에 따라 서로 다른 오차를 보이지만, iteration이 반복될수록 복원 오차는 점차 감소한다. 이러한 경향은 기존 보간 기법으로 초기값을 설정한 경우뿐 아니라, 초기값을 무작위 상수로 부여한 경우에도 동일하게 나타났으며, 이를 통해 iteration이 충분히 진행되면 초기 보간 기법 및 초기값의 설정과 무관하게 복원값이 하나의 안정적인 값으로 수렴하는 모습을 확인할 수 있다. 이는 표 1~3에서 동일한 결측 비율에 대해 제안 알고리즘을 적용한 보간 기법들의 MAE와 RMSE 성능이 서로 유사한 값을 보이는 결과와도 일치한다.

V. 결 론

본 논문에서는 시계열 데이터의 결측치 복원에서 기저 보간 기법의 선택에 따라 복원 성능 편차가 발생하는 문제를 완화하고자, 반복적 이웃 평활화 기반 보간 알고리즘인 INSI를 제안하였다. 제안 알고리즘은 결측 시점 주변 이웃값의 거리 기반 가중 평균을 반복적으로 반영하여 보간값을 업데이트하며, 수학적으로 초기 보간값과 무관하게 하나의 유일한 값으로 수렴함을 증명하였다. 제안 알고리즘의 성능을 검증하기 위해 선형 보간, KNN, ARIMA, BRITS, SAITS를 활용하여 단일 보간 기법과 제안 알고리즘이 후처리 보간 방식으로 적용되었을 때의 복원 성능을 비교하였으며, 실험 결과 기존 알고리즘 대비 제안 알고리즘의 MAE와 RMSE가 각각 평균 약 68% 향상되었다. 더불어, 결측치를 임의의 무작위 상수값으로 대체하고 제안 알고리즘과 기존 보간 기법들의 성능을 비교한 단일 보간 실험에서도 모든 데이터셋에서 안정적인 복원 성능을 보였으

며, 이를 통해 제안 알고리즘이 단독 보간 알고리즘으로도 활용될 수 있음을 확인하였다. 이러한 결과는 제안 알고리즘이 기저 보간 기법이나 초기값 설정과 무관하게 안정적인 복원 성능을 제공할 수 있음을 보여주며, 향후 다양한 산업 환경의 시계열 데이터 결측치 복원 문제에 활용될 수 있을 것으로 기대된다.

참 고 문 헌 (References)

[1] S. S. W. Fatima and A. Rahimi, “A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems,” Machines, Vol. 12, No. 6, Article 380, June 2024.
doi: https://doi.org/10.3390/machines12060380

[2] J. Du, M. Hu, and W. Zhang, “Missing Data Problem in the Monitoring System: A Review,” IEEE Sensors Journal, Vol. 20, No. 23, pp. 13984-13998, Dec. 2020.
doi: https://doi.org/10.1109/JSEN.2020.3009265

[3] G. M. Phillips, Interpolation and Approximation by Polynomials, Springer, New York, NY, USA, 2003.
doi: https://doi.org/10.1007/b97417

[4] F. N. Fritsch and R. E. Carlson, “Monotone Piecewise Cubic Interpolation,” SIAM Journal on Numerical Analysis, Vol. 17, No. 2, pp. 238-246, Apr. 1980.
doi: https://doi.org/10.1137/0717021

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed., John Wiley & Sons, Hoboken, NJ, USA, 2015. Available: Wiley online page.

[6] O. G. Troyanskaya, M. N. Cantor, G. Sherlock, P. O. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, “Missing Value Estimation Methods for DNA Microarrays,” Bioinformatics, Vol. 17, No. 6, pp. 520-525, June 2001.
doi: https://doi.org/10.1093/bioinformatics/17.6.520

[7] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, “BRITS: Bidirectional Recurrent Imputation for Time Series,” in Advances in Neural Information Processing Systems 31, Montréal, Canada, pp. 6776-6786, 2018.
doi: https://doi.org/10.5555/3327757.3327783

[8] W. Du, D. Côté, and Y. Liu, “SAITS: Self-Attention-Based Imputation for Time Series,” Expert Systems with Applications, Vol. 219, Article 119619, June 2023.
doi: https://doi.org/10.1016/j.eswa.2023.119619

[9] S. M. Ribeiro and C. L. de Castro, “Missing Data in Time Series: A Review of Imputation Methods and Case Study,” Learning and Nonlinear Models, Vol. 20, No. 1, pp. 31-46, Oct. 2022.
doi: https://doi.org/10.21528/lnlm-vol20-no1-art3

[10] Y. Bengio, P. Simard, and P. Frasconi, “Learning Long-Term Dependencies with Gradient Descent is Difficult,” IEEE Transactions on Neural Networks, Vol. 5, No. 2, pp. 157-166, Mar. 1994.
doi: https://doi.org/10.1109/72.279181

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, pp. 5998-6008, 2017.
doi: https://doi.org/10.5555/3295222.3295349

[12] J. Yoon, J. Jordon, and M. van der Schaar, “GAIN: Missing Data Imputation Using Generative Adversarial Nets,” in Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR, Vol. 80, pp. 5689-5698, 2018. Available: PMLR page.

[13] Y. Tashiro, J. Song, Y. Song, and S. Ermon, “CSDI: Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation,” in Advances in Neural Information Processing Systems 34, pp. 24804-24816, 2021.
doi: https://doi.org/10.5555/3540261.3542161

저 자 소 개

김파란하늘

- 2023년 3월 ~ 현재 : 국립한밭대학교 인공지능소프트웨어학과 학사과정
- ORCID : https://orcid.org/0009-0005-0421-238X
- 주관심분야 : 시계열 데이터 분석, 컴퓨터 비전, 대규모 언어모델(LLM)

저 자 소 개

이 연 서

- 2023년 3월 ~ 현재 : 국립한밭대학교 인공지능소프트웨어학과 재학
- ORCID : https://orcid.org/0009-0007-0191-8111
- 주관심분야 : 시계열 데이터 분석, 컴퓨터 비전

현 장 훈

- 2014년 : 고려대학교 전기전자전파공학부 공학사
- 2021년 : 고려대학교 전기전자공학부 공학박사
- 2021년 ~ 2023년 : 고려대학교 연구교수
- 2023년 ~ 현재 : 국립한밭대학교 인공지능소프트웨어학과 조교수
- ORCID : https://orcid.org/0000-0001-5737-9179
- 주관심분야 : 컴퓨터 비전, 로보틱스, 딥러닝, 센서퓨전