

제철 공정의 용강 온도 예측을 위한 전로 연속성 기반 듀얼 어텐션 순환 신경망 모델

한경석 · 김성범[†]

고려대학교 산업경영공학과

Dual Attention Recurrent Neural Network for Molten Steel Temperature Prediction with Converter Continuity

Kyung Seok Han · Seoung Bum Kim

School of Industrial and Management Engineering, Korea University

The steelmaking process involves the removal of impurities from molten iron to achieve the desired material properties and ensure the quality of the final steel product. Proper control of molten steel temperature is crucial because it directly and indirectly impacts product quality. Inadequate temperature control can lead to defects in the final product or failure to meet the desired material specification. Traditionally, machine learning algorithms have been applied to predict molten steel temperature by treating each steelmaking cycle in the converter as an independent event. However, such simple regression-based approaches often fail to capture the sequential characteristics of the converter process. In this study, we propose a time-series modeling approach that incorporates residual heat information from previous cycles to better reflect the continuous operational characteristics of the converter. This method enables the identification of operational patterns that conventional models overlook and demonstrates improved temperature prediction accuracy. The proposed predictive model is expected to contribute to future optimization of raw material and energy inputs required to achieve the target temperature.

Keywords: Time-Series Modeling, Regression, Neural Network, Steelmaking, Converter

1. 서론

제철 공정은 원료의 환원부터 최종 제품의 압연에 이르기까지 일련의 연속된 공정으로 이루어진다. 일반적으로 제철 공정은 <Figure 1>과 같이 제선, 제강, 연주, 압연의 네 단계로 구분된다. 제선(製銑, ironmaking) 공정은 철광석을 녹여 쇳물을 만드는 공정이다. 제철 원료인 철광석과 석탄은 각각 소결 공장과 코크스 공장을 통해 소결광, 코크스로 만들어지고, 이들을 고로(高爐, blast furnace) 위에서 장입한다. 고로 아래에서는 1,200°C의 뜨거운 바람을 불어넣어 코크스를 산화시키고, 동시에 소결광을 환원시켜 용선(molten iron)을 만든다. 제강(製

鋼, steelmaking) 공정은 용선에 포함된 불순 원소(C, Si, Mn, P 등)를 제거하여 깨끗한 상태의 쇳물인 용강(molten steel)을 제조하는 단계이다. 대표적인 방식으로는 전로(轉爐, converter), 전기로(電氣爐, electric arc furnace)를 이용한 제강이 있으며, 본 논문에서 다룬 전로 공정에서는 고순도 산소를 불어넣어 산화 반응을 유도함으로써 불순물을 제거한다. 연주(連鑄, continuous casting) 공정은 연속 주조 공정의 줄임말로, 액체 상태의 용강을 응고시켜 고체 상태의 슬래브(slab), 블룸(bloom), 빌렛(billet) 등 중간재 형태로 제조하는 단계이다. 응고 속도, 주형의 냉각 조건 등이 제품 품질에 직접적인 영향을 미친다. 압연(壓延, rolling) 공정은 연주 공정에서 제조된 반제품을 다

[†] 연락저자 : 김성범 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학부, Tel: 02-3290-3397, Fax: 02-929-5888,
E-mail : sbkim1@korea.ac.kr

2025년 9월 4일 접수; 2025년 11월 3일 수정본 접수; 2025년 11월 13일 게재 확정.

양한 크기와 두께의 철강제품으로 만드는 공정이다. 슬래브를 1,100°C 이상 가열 후 대형 압연롤을 통과시키면 열연제품이 되고, 두꺼운 철판으로 자르면 후판제품이 된다. 빌렛은 긴 철선의 선재제품으로 가공된다.

본 연구에서 다루는 제강 전로 공정은 제철 공정 중 유일하게 액체 상태의 철을 제어하는 공정으로, 화학 조성을 조정하고 쇳물의 청정도를 확보함으로써 최종 제품의 강도, 연신율 등의 기계적 특성과 품질에 직접적인 영향을 미치는 핵심 역할을 담당한다. 전로 공정은 <Figure 2>와 같이 장입(charging), 취련(blowing), 판정(judgement), 출강(tapping)의 네 단계로 구성된다.

첫 번째 단계인 장입은 스크랩과 용선을 전로에 장입하고, 원소의 함량과 온도 등의 정보를 기반으로 취련을 준비하는 과정이다. 두 번째 단계인 취련은 본 연구의 핵심으로, 뒤에 자세히 설명하도록 하겠다. 세 번째 단계인 판정은 취련 완료 시점의 온도, 즉 중점 온도를 측정하고 후속 조치 여부를 결정하는 단계이며, 마지막 출강 단계에서는 취련이 완료된 용강을 ‘래들’ 이라 불리는 용기에 따라내어 후속 공정으로 이송한다. 이와 같이 네 단계를 마치면 하나의 전로 작업이 완료되며, 이를 한 개의 ‘차지(charge)’라 한다. 차지 단위 작업은 반복적으로 수행되며, 하나가 종료되면 곧바로 다음 스크랩과 용선이 장입된다.

전로 공정의 핵심인 취련 단계에서는 순도 99% 이상의 산소를 랜스 노즐을 통해 용선에 초음속으로 취입(불어넣음)하여, 각 원소의 산화 반응을 유도한다. 이는 곧 산소를 불어넣어 용선을 정련하는 과정이며, 탄소 함량이 약 2% 이하로 감소하면 용선은 용강으로 전환된다. 이 때 탄소를 비롯한 각 원소는 아래의 반응식을 통해 산화된다.



위 반응들은 모두 발열 반응으로, 용선 혹은 용강의 온도를 자연스럽게 상승시킨다. 이러한 온도 상승은 원소의 산화 반응에 따른 불가피한 물리화학적 현상인 동시에, 목표 온도를 확보하기 위한 필수적인 열원이기도 하다. 여기서 말하는 ‘목표 온도’란, 후속 공정 조건, 처리 시간, 제품 사양 등에 따라 사전에 설정되는 온도 기준을 의미한다. 전로에서는 용강의 온도를 목표 온도까지 상승시키기 위하여 산화 반응열을 적극 활용하되, 필요한 경우 추가 열원의 투입이나 냉각재 투입 등 보조 조치를 통해 온도를 조정한다. 그러나 이 과정이 실패하

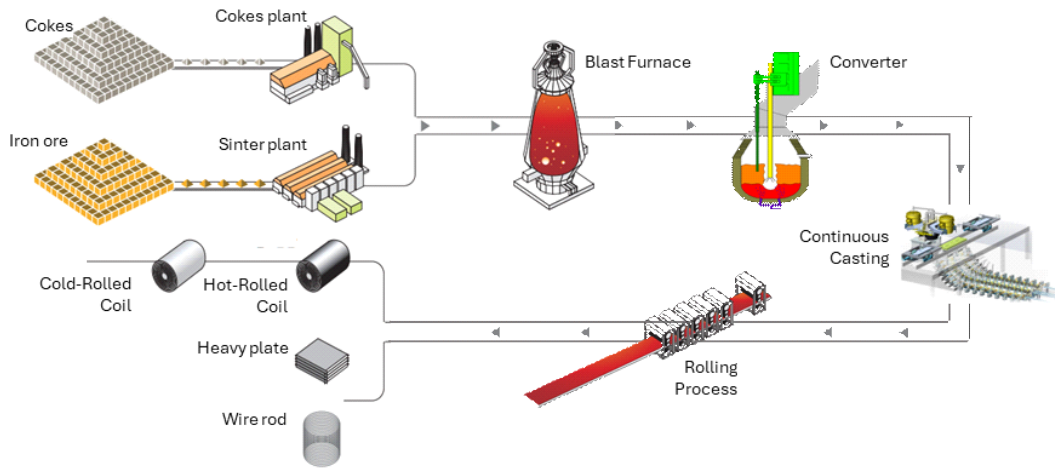


Figure 1. Overview of the Steel Manufacturing Process: Raw Materials, Such as Iron Ore and Cokes go Through the Blast Furnace, Converter, and Continuous Casting to Produce Semi-finished Products, Which are Then Rolled Into Final Products

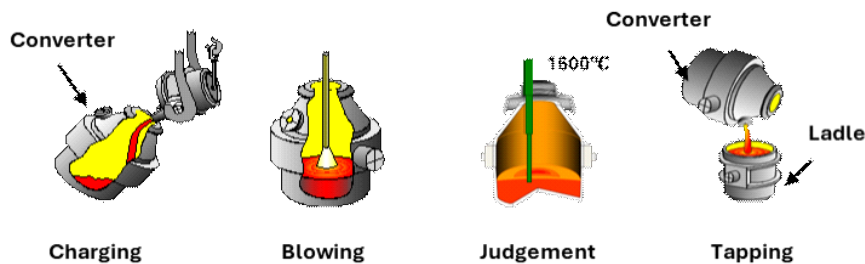


Figure 2. Diagram of the Converter Process Including Charging, Blowing, Judgement and Tapping

여 목표 온도를 $\pm 10^{\circ}\text{C}$ 이상 벗어날 경우, 후속 공정과 제품 품질에 크고 작은 문제가 발생할 수 있다. 예를 들어 온도가 과도하게 높아지는 경우, 이는 불순 원소의 산화가 완료된 이후에도 산소가 과도하게 투입되어 용철(liquid iron) 자체가 산화되었음을 의미하며, 이로 인해 제품 실수율이 하락하고, 용강 내 과도한 활성 산소(다른 원소와 반응하지 않고 남아 있는 산소)가 후속 공정의 응고 과정 및 최종 제품 품질에 악영향을 끼친다. 반대로 온도가 낮은 경우, 이는 불순 원소들이 충분히 제거되지 않았음을 의미하며 이러한 경우 원하는 기계적 특성 확보가 어렵고, 온도가 부족하므로 후속 공정 수행에도 문제가 발생한다. 특히 온도를 보충하기 위한 추가 작업은 품질 저하를 야기하는 주요 원인이 된다.

따라서 용강 온도 제어는 후속 공정을 원활하게 하고 최종 제품의 품질을 확보하는 매우 중요한 과정이다. 그러나 실제 용강의 온도 거동은 단순한 산화 반응의 열량 계산만으로는 설명할 수 없는 복합적인 물리화학적 상호작용의 결과이며, 이를 이론적으로 완벽히 정립하기란 불가능하다. 그 주요한 이유는 다음과 같다. 첫째, 수 십여 종의 스크랩이 용선을 얼마나 냉각시키는지에 대한 이론적, 실험적 결과가 부족하다. 둘째, 쇳물(용선 혹은 용강)이 전로 내부의 내화 벽돌과 접촉하며 손실되는 열량을 정량화하기 어렵다. 셋째, 용선에 투입되는 각종 원료들이 가져가는 열량을 정확히 계산하기 어렵다. 넷째, 전로 자체가 보유한 축열량(蓄熱量, 머금고 있는 열)을 정밀하게 계산할 수 없다. 마지막으로 산소의 반응 효율을 실시간으로 확인할 방법이 없다.

현업에서는 위와 같이 알 수 없는 요소들을 배제하고, 알고 있는 것에 집중하여 이론적 계산을 수행하지만, 이러한 계산은 상당한 숙련을 요구하고, 계산 방법이 작업자마다 각자의 편리대로 달라지며, 결정적으로 제대로 계산하더라도 그 결과가 정확하지 않다. 이러한 문제를 해결하기 위하여, 온도 거동을 예측하는 다양한 자동화 시도가 오랜 시간동안 이루어져 왔으나, 앞서 언급한 여러 한계로 인하여 만족할 만한 성과를 얻지 못하였다. 특히, 출강 이후 전로 내부에 남은 잔류물의 열량과 전로 자체의 축열량은 다음 차지의 온도 거동에 중대한 영향을 미치지만, 이를 정량적으로 파악할 수 있는 센서나 데이터가 존재하지 않아 정확한 온도를 예측하는 데 어려움이 있었다. 이는 온도 거동에 실질적인 영향을 미치는 물리적 요소들이 분명 존재함에도 불구하고, 이를 실질적으로 활용할 수 없는 구조적인 한계에서 비롯된 문제라 할 수 있다.

이에 본 연구에서는 잔류물의 열량 정보를 명시적으로 입력하지 않고도, 모델이 이를 간접적으로 학습할 수 있도록 시계열 회귀 기반의 예측 모델을 설계하였다. 즉, 전로 공정의 각 차지를 독립적인 작업으로 가정하지 않고, 시간의 흐름에 따른 연속적인 시계열 시스템으로 해석하여, 앞선 차지의 축열량을 모델이 간접적으로 학습하고, 후속 차지에서 이를 활용할 수 있도록 하여, 실제 조업에서의 물리적인 작동 방식을 그대로 반영한 것이다. 이를 통해 기존 단순 회귀 기반 접근법이

설명하지 못하던 온도 변화 패턴을 포착할 수 있었고, 온도 예측 성능을 유의미하게 향상시켰으며, 이를 바탕으로 향후 목표 온도 달성을 위한 원료 투입량 및 산소 제어의 최적화를 기대할 수 있다.

2. 관련 연구

최근 철강 산업은 글로벌 공급 과잉으로 인한 경쟁 심화와 더불어, 기후 변화 대응을 위한 국제 환경 규제 및 탄소 중립 요구에 따라 저탄소, 고품질 제품 생산을 동시에 달성해야 하는 압박을 받고 있다(Li *et al.*, 2023). 이러한 상황에서, 생산 공정의 자동화를 통해 인적 편차를 최소화하고 공정의 일관성과 효율을 높이기 위해 인공지능 기술의 도입이 적극적으로 시도되고 있다. 특히 제강 공정은 고온의 환경에서 다양하고 복잡한 변수들이 물리화학적으로 상호작용하는 시스템으로, 기존 경험 기반 운영이나 단순 수식 기반 모델만으로는 공정의 복잡한 거동을 정확히 모사하는데 한계가 있다. 이에 따라 인공지능 기반 예측 및 제어 기술은 제강 공정의 자동화를 실현하기 위한 유력한 수단으로 주목받고 있다. 그러한 시도의 일환으로 제강 공정의 온도 예측을 목표로 한 다양한 인공지능 기반 연구들이 보고되고 있다.

Lee *et al.*(2019)은 제강 공장의 조업 데이터를 바탕으로 XGBoost, LightGBM, CatBoost 기반의 예측 모델과 이를 조합한 mixture of experts 앙상블 모델을 제안하였다. Feng *et al.*(2021)은 전로의 종점 온도를 예측하기 위해 support vector regression(SVR), random forest(RF), backpropagation(BP) 신경망 세 가지 모델을 결합한 베이지안 가중 앙상블 모델을 제안하였다. 이들은 각 모델의 예측 성능에 따라 동적으로 가중치를 조정하는 방식을 통해 모델을 통합하였고, root mean squared error (RMSE) 3.86°C 의 예측 정확도를 보였다. Gu *et al.*(2021)은 전로의 종점 온도와 탄소 함량을 예측하기 위해, 사례 기반 추론(case-based reasoning, CBR)에 시계열 데이터를 통합한 모델을 제안하였다. CBR은 과거 유사한 사례로부터 새로운 문제의 해답을 유추하는 방식으로, 과거 입력 값 중 현재 입력 조건과 유사한 사례를 찾고 해당 사례의 결과를 근거로 예측을 수행하는 방식이다. 이들은 입력 변수를 정적 데이터와 시계열 데이터로 구분하고, 산소 유량, 랜스 위치, 저취 가스 유량과 같은 시계열 변수에 대해 dynamic time warping을 사용하여 시계열간 유사도를 계산하였다. 이후 정적 변수 유사도와 시계열 유사도를 가중 평균하여 최종 유사도를 산출하고, 이 유사도를 기반으로 선택된 사례들의 결과값을 평균 또는 가중 평균하여 예측값을 도출하였다. Qiu *et al.*(2024)은 전로 공정에서의 종점 온도 및 탄소 함량 예측 정확도를 개선하기 위해, 이상치 제거 및 계층적 클러스터링, principal component analysis(PCA) 기반 차원 축소, 무작위 샘플링 보정 등 다양한 전처리 기법을 적용한 후, chaotic sparrow search 알고리즘으로 최

적화된 BP 신경망을 설계하였다. 그 결과, 비교 모델 중 가장 낮은 mean absolute error(MAE) 및 RMSE를 기록하며 우수한 예측 성능을 보였고, Shapley additive explanation(SHAP)을 활용하여 모델 해석 가능성도 함께 제시하였다.

이상과 같이, 기존 연구들은 다양한 회귀 기반 모델과 신경망 구조를 활용하여 전로 종점 온도 예측의 정확도를 높이기 위한 시도를 이어왔으며, 일부 연구에서는 시계열 정보를 부분적으로 활용하거나 고차원 데이터를 정교하게 가공하여 성능을 개선하기도 하였다. 그러나 이들 대부분은 여전히 각 차지를 독립적인 조업 단위로 간주하고 있으며, 전로 내부에 누적되는 잔열이나 축열과 같은 물리적 연속성을 모델에 반영하지 못한다는 공통된 한계를 지닌다. 실제 조업 현장에서는 하나의 차지가 끝난 이후에도 전로 내부에 열과 잔류물이 잔존하게 되며, 이는 다음 차지의 온도 거동에 직간접적으로 영향을 미친다. 이에 본 연구에서는 전로 조업 전체를 시계열 데이터로 간주하고, 조업 간 연속적인 열적 영향을 모델이 학습하도록 유도하는 구조를 제안한다. 이를 위해 시계열 데이터 처리에 효과적인 구조인 gated recurrent unit(GRU) 기반의 회귀 모델을 적용하였다. GRU는 Cho *et al.*(2014)에 의해 제안된 구조로, 입력 시퀀스의 시간적 패턴을 효율적으로 학습하고, 이전 시점의 정보를 은닉 상태에 저장하여 다음 시점으로 전달함으로써 장기 의존성 문제를 완화하는 순환 신경망 모델이다. 이러한 접근을 통해 본 연구는 기존의 정적 회귀 방식으로는 반영할 수 없었던 전로 내부 상태를 포착하도록 하고, 온도 변화 패턴을 보다 정밀하게 예측하고자 한다.

3. 제안 방법론

3.1 데이터 수집 및 전처리

(1) 데이터 수집

데이터는 대한민국 광양제철소 3제강공장의 데이터를 사용

하였다. 샘플 수는 총 35,440차지로, 전로 및 후속 공정, 그리고 최종 목표 변수인 ‘종점 온도’ 데이터를 포함하여 총 102개의 변수로 구성되어 있으며, 이 중 전로 공정과 직간접적으로 연관이 있는 24개의 독립변수를 선정하였다. 또한 용선 중 탄소, 실리콘 등과 같은 불순 원소가 완전히 산화되었을 때의 이론 발열량을 계산하고 이를 용강 온도로 환산하여 별도의 독립변수로 지정하였다.

(2) 연속 조업 사이클

전로 조업은 각 차지가 별도의 용선과 스크랩으로 운영되는 독립된 작업의 반복으로 보이지만, 각 차지의 출강이 완료된 이후 남아 있는 잔류 용강 및 슬래그(취련 과정 중 발생하는 복합 산화물)를 기술적으로 활용하기 위하여 완전히 버리지 않고 일부를 남긴 상태에서 다음 용선을 전로에 장입 한다. 따라서 전로 조업은 독립된 작업의 반복이 아니라, 물리적·열역학적으로 연속된 흐름을 가지는 공정으로 해석되어야 한다. 다만, 잔류 슬래그는 전로의 가동 정지 전까지 계속 유지되는 것은 아니며, 설비 관리 목적이나 특정한 공정 조건에 따라 완전히 제거되는 경우도 존재한다. 이러한 제거는 일정한 주기를 따르지는 않지만, 조업 현장에서 비교적 빈번하게 발생하는 현상이다.

본 연구에서는 전로 조업의 시계열적 특성, 즉 이전 차지에서 남겨진 잔류 슬래그가 다음 차지에 영향을 미치는 연속성을 모델에 반영하고자 하였다. 이를 위해 조업 간 연속 여부를 잔류 슬래그의 존재 유무를 기준으로 판단하였다. 예를 들어, 특정 차지에서 잔류 슬래그가 존재하는 경우 이를 ‘1(있음)’으로 표기하며, 해당 차지는 이전 차지로부터 연속된 조업의 일부로 간주된다. 반대로, 잔류 슬래그가 존재하지 않는 경우에는 이를 ‘0(없음)’으로 표기하고, 해당 차지는 새로운 조업 사이클의 시작점으로 정의한다. 즉, 잔류 슬래그가 존재하면 연속 조업(cycle) 내에 있는 차지로 간주되며, 존재하지 않으면 새로운 조업 사이클이 시작된 것으로 해석한다. 이러한 개념을 <Figure 3>에 도식화 하였다.

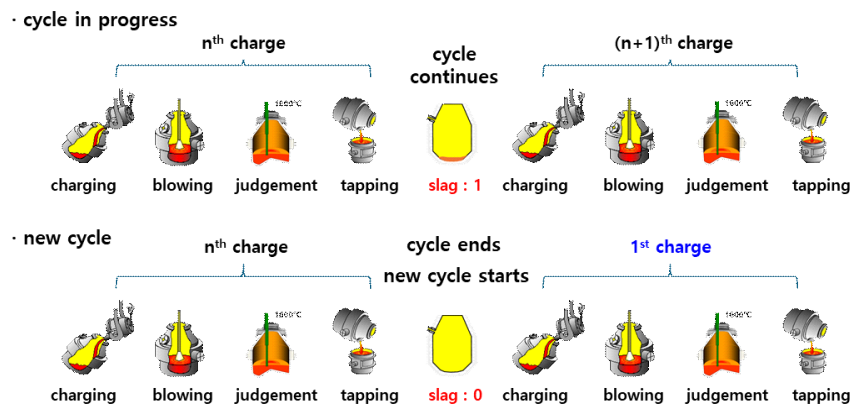


Figure 3. Schematic Illustration of Converter Cycle Continuity Determination: The Cycle is Considered Continuous when Residual Slag from the Previous Charge Is Carried over (top). If no residual slag remains, the current cycle ends and a new one begins (bottom).

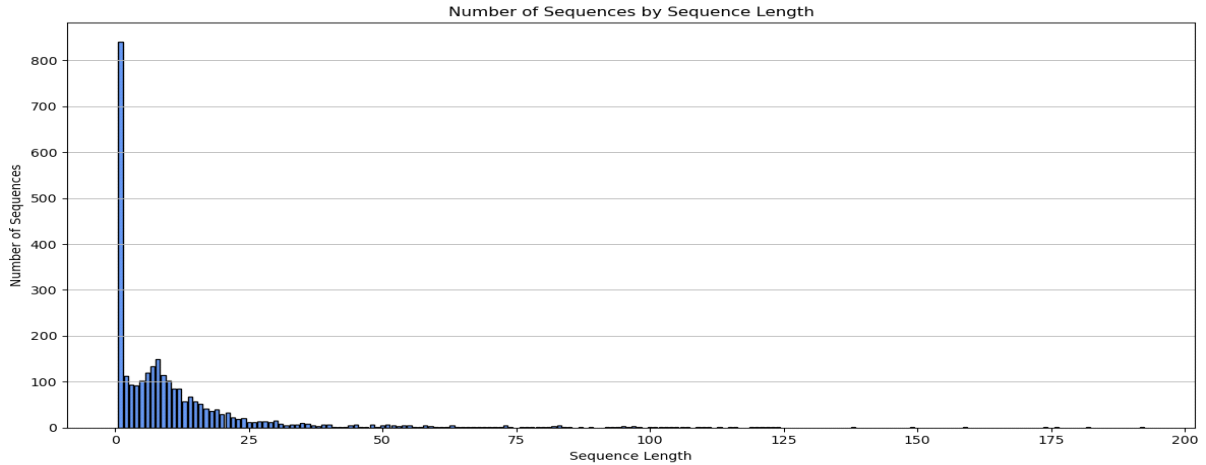


Figure 4. Frequency Distribution of Converter Cycle Lengths: The Histogram Shows that Most Cycles are Short, with 95% Lasting 50 units or Less

위와 같은 방법으로 총 35,440개의 차지를 연속 조업 사이클로 묶은 결과, 총 2,806개의 사이클이 구성되었다. 이 중 가장 긴 사이클은 192차지에 달하며, 가장 짧은 사이클은 1차지로 구성되었다. 사이클 길이에 따른 분포는 <Figure 4>에 제시하였다.

분포를 분석한 결과, 전체 사이클의 약 95%가 50차지 이하의 길이를 가지는 것으로 나타났다. 이에 따라, 50차지를 초과하는 장주기 사이클에 대해서는 앞부분의 50차지만을 남기고 나머지는 제거하였다. 이는 장주기 사이클의 발생 빈도가 매우 낮아 학습에 유의미한 기여를 하기 어려울 뿐 아니라, 장주기 데이터 처리에 소요되는 자원을 최소화하기 위한 조치이다. 또한 실제 조업 현장에서도 이와 같은 장주기 연속 조업은 드물게 발생하므로, 본 연구에서는 보다 일반적인 조업 패턴에 초점을 맞추고자 하였다. 다만, 이러한 절단 방식이 후반부의 조업 패턴을 학습에서 배제함으로써 잠재적인 성능 편향, 즉 50차지 이내의 ‘짧은’ 사이클에만 과적합 될 수 있다는 가능성을 고려할 필요가 있다. 그러나 본 연구에서는 이러한 편향이 모델의 전반적인 예측 성능에 미치는 영향은 미미할 것으로 판단하였다. 그 이유는 첫째, 사이클 전반에 걸친 온도 변화에 결정적인 영향을 미치는 정보의 중요도는 <Figure 15>에서 확인할 수 있듯이 사이클 초반부에 집중된다. 이는 모델이 50차지 이내의 데이터만으로도 공정의 핵심적인 물리화학적 패턴을 학습하기에 충분함을 시사한다. 둘째, 본 연구의 목적은 50 스텝 이후 장주기 사이클의 후반부 예측이 아니라 각 시점에서의 온도 예측이다. <Figure 4>에서 확인할 수 있듯이 50차지를 초과하는 5% 미만의 데이터는 전체 조업의 일반성을 대표하기 어려우므로, 95%의 일반적인 조업 패턴에 집중하는 것이 모델의 전반적인 일반화 성능을 확보하는 데 더 합리적이다. 또한 실제로 50차지를 초과하는 경우가 발생하더라도, 모델은 이미 50차지 이내에서 학습한 핵심 패턴을 기반으로 해당 시점의 입력을 처리하여 합리적인 예측을 수행할 수 있다.

전처리가 완료된 전체 2,806개의 조업 사이클은 훈련, 검증, 테스트 셋으로 70%, 15%, 15%의 비율로 분할하였다. 일반적인 시계열 분석에서는 정보 누수를 방지하기 위해 데이터를 시간 순서대로 분할한다. 그러나 본 연구에서는 의도적으로 무작위 분할 방식을 채택하였으며, 이는 후술할 제강공정의 전로 수명을 고려한 것이다. 전로 수명은 일반적인 시간의 흐름처럼 한방향으로 무한히 증가하는 변수가 아니라, 3.1.3절에서 설명하는 바와 같이 수 천 회의 조업 후 내화 벽돌을 전면 교체하면서 다시 1회부터 시작하는 주기적 상태 변수이다. 따라서 1년차와 3년차의 전로 수명 n 회는 달력상의 시간은 다르지만 전로의 물리적 상태는 동일한 분포에서 추출된 표본으로 간주할 수 있다. 따라서 이와 같은 무작위 분할 방식은, 특정 연도나 특정 전로 일대기에만 국한된 편향을 방지하기 위한 것이다. 즉, 모델이 물리적 시간이 아닌 ‘전로 수명’이라는 특수한 시간의 흐름에 일반화된 조업 패턴을 학습하도록 유도하는 더 강건한 설계이다.

(3) 전로 수명

3.1.2절의 연속 조업 사이클 구성을 바탕으로, 각 차지는 자신이 속한 사이클 내에서 몇 번째 차지인지를 나타내는 상대적인 순서 정보(상대 수명)를 가지게 된다. 동시에, 각 차지는 전로 가동 개시 이후 전체 작업 흐름에서 몇 번째 차지인지를 나타내는 절대적인 순서 정보(절대 수명)도 함께 갖는다. 이 두 순서 정보는 비슷해 보이지만, 각각 서로 다른 조업 특성을 반영하고 있다. 우선, 상대 수명은 각 차지가 속한 사이클 내에서의 위치를 나타내며, 앞서 언급한 잔류 슬래그의 열량 및 전로 자체의 축열량 등 조업 간 열적 연속성과 관련된 정보를 내포한다. 이는 차지 간 물리적 연결성과 축열량을 모델이 학습하는 데 중요한 역할을 한다. 반면, 절대 수명은 전로 가동된 이후 해당 차지가 몇 번째 작업인지를 나타내는 1씩 증가하는 단순 누적 순서이다. 그러나 이는 단순한 순번 정보에 그치

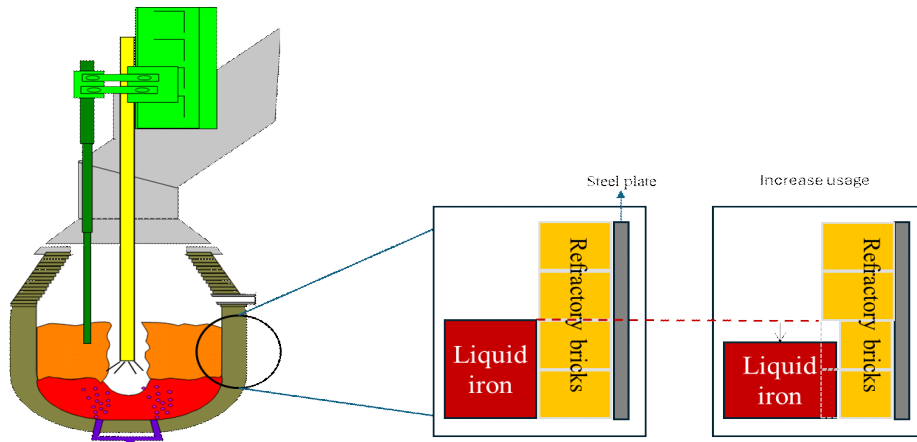


Figure 5. Schematic Illustration of Refractory Brick Erosion: Erosion Expands the Converter’s Internal Volume, Leading to a Decrease in the Liquid Iron Height

지 않는다. 전로 공정이 반복될수록 고온·고압 환경에 의한 내화 벽돌의 침식이 발생하고, 이로 인해 전로의 내용적 (volume)이 점차 증가하게 된다(<Figure 5>). 이러한 구조 변화는 산소 반응 효율, 용강의 유동 특성 등 공정 전반에 물리화학적 불확실성을 초래하며, 조업 품질 및 예측 정확도에도 영향을 미친다. 따라서 절대 수명은 단순한 시간 흐름 이상의 의미를 가지며, 조업 전반의 특성과 장기적인 변화 추이를 간접적으로 반영하는 지표이다.

전로의 최대 절대 수명은 제강 공장마다 상이하며, 현장 조건 및 설비 운용 상황에 따라 유동적으로 결정된다. 일반적으로 전로는 수천 차지의 작업을 수행한 후 가동을 중단하고, 내화 벽돌을 전면 교체하는 과정을 거친다. 예를 들어, A 제강 공장의 평균 전로 수명이 5,000차지이고, 하루 평균 30차지의 조업이 이루어진다고 가정하면, 약 167일마다 전로가 정지되며, 산술적으로 연간 2.18회의 가동 중단이 발생하게 된다. 이는 곧, 절대 수명 1부터 5,000까지 각각의 수명 구간에서의 조업 데이터가 연간 최대 2.18회만 관측될 수 있다는 뜻이며, 절대 수명에 따른 조업 특성을 충분히 학습하기에는 데이터의 양이 절대적으로 부족함을 의미한다. 이와 같은 데이터 희소성은 절대 수명이 조업 거동에 미치는 잠재적 영향을 효과적으로 학습하는 데 있어 중요한 제약 요소로 작용한다. 이러한 문제를 완화하기 위해, 본 연구에서는 전로 수명을 다음과 같이 구간화 하여 조업 단계를 정의하였다.

- 절대 수명 1~50회: 전로 가동 초기의 시작 단계’
- 절대 수명 51~500회: 조업 조건이 비교적 안정화된 초기 단계’
- 절대 수명 501~3,000회: 전형적인 조업이 수행되는 중기 단계’
- 절대 수명 3,001회 이상: 내화 벽돌의 열화와 구조적 변형이 본격화되는 말기 단계’

이와 같은 구간화는 단순히 수치를 기준으로 분할한 것이

아니라, 실제 조업 경험과 전로 운영의 물리적 특성을 반영하여 설계된 것이다. 이를 통해 모델이 전로의 수명에 따른 구조적 변화와 조업 조건의 차이를 보다 효과적으로 학습할 수 있도록 하였다. 이상과 같은 정보를 종합하면, 각 차지는 <Table 1>의 예시와 같이 잔류 슬래그의 유무, 상대 수명, 절대 수명, 그리고 조업 단계의 데이터를 포함한다.

Table 1. Example Data Showing Residual Slag Presence, Relative Life, and Absolute Life: Relative Life Counts Charges within a Continuous Cycle and Resets to 1 when Residual Slag is Absent, and Absolute Life Counts Total Charges to Indicate the Process Stage

Charge index	Residual slag	Relative life	Absolute life	Process stage
1	1	21	499	Early
2	1	22	500	Early
3	0	1	501	Mid
4	1	2	502	Mid
5	1	3	503	Mid

(4) 결측치, 이상치 처리 및 정규화

위와 같이 수집된 데이터에 대하여 다음과 같은 전처리 절차를 수행하였다.

첫째, 본 연구는 전로 조업의 시계열 연속성을 전제로 하므로, 센서 오류나 수집 누락으로 인한 결측치가 발생한 차지를 단순히 제거할 경우 사이클의 순차적 구조가 파괴되는 문제가 발생한다. 이러한 교란을 방지하고 데이터의 연속성을 보존하기 위해, 누락된 값은 선형 보간법(linear interpolation)을 사용하여 대체하였다.

둘째, 이상치(outlier)는 통계적 기법을 적용하는 대신, 실제 조업 환경과 공정의 물리적 한계를 고려한 도메인 기반의 필터링을 적용하였다. 예를 들어, 총 산소 투입량이 정상 조업 범

위를 (e.g., 11,500~16,000Nm³) 벗어나거나, 중점 온도가 통상의 범위(e.g., 1600~1750℃)를 벗어나는 경우, 분석에서 제외하였다. 이러한 비정상 데이터의 수는 전체 데이터의 0.5% 수준으로, 이로 인해 발생하는 일부 시퀀스의 연속성 단절은 감안할 수 있는 수준으로 판단하였다.

셋째, 데이터 정규화는 입력 변수와 목표 변수에 각기 다른 방식을 적용하였다. 입력 변수는 모든 변수가 평균 0, 분산 1을 갖도록 standard scaler를 적용하였다. 이는 모델이 모든 변수의 스케일에 구애 받지 않고 동등한 중요도로 학습하도록 돕는다. 목표 변수인 ‘중점 온도’는 0과 1 사이의 값을 갖도록 minmax scaler를 적용하였다. 이 때, 모든 scaler는 훈련 데이터셋에 대해서만 fitting하여 데이터 누수(data leakage)를 방지하였다.

3.2 제안 방법론

(1) 독립 변수 그룹핑 및 임베딩

본 연구에서는 모든 입력 변수를 단순히 하나의 벡터로 통합하여 사용하는 기존 방식에서 벗어나, 각 변수의 물리적 의미와 조업 과정에서의 데이터 특성을 기준으로 변수들을 그룹화하였다. 예를 들어, 용선의 화학적 성분인 카본, 실리콘 등의 함량과 용선 온도를 하나의 카테고리로 묶고, 스크랩 종류별

투입량에 해당하는 12개의 변수는 또 다른 카테고리로 분류하였다. 이와 같은 접근법은 계강 조업 데이터의 특성을 고려한 것으로, 예를 들어 12개의 변수로 구성된 ‘스크랩 종류별 투입량’은 쇳물의 열을 빼앗아 간다는 공통된 특성을 가지고 있고, ‘용선 정보’ 카테고리는 온도, 성분 함량, 무게 등 서로 다른 스케일과 물리량을 가지지만, 결국 용선이 가지고 있는 열량을 산출하기 위한 변수들이라는 공통점이 있다. 이러한 카테고리별 특성은 <Table 2> 와 같다.

본 연구에서 제안하는 카테고리 기반의 계층적 입력 구조는 조업 데이터가 가진 특성을 효과적으로 처리하기 위해서 설계되었다. 일반적인 회귀 분석에서처럼 모든 변수를 단일 벡터로 처리할 경우 두 가지 문제가 발생할 수 있다. 첫 번째는 ‘Molten iron’ 그룹과 같이 온도, 무게와 같은 이질적인 변수들로 이루어져 있지만 결국에는 용선의 열량으로 귀결된다는 내재적 관계를 모델이 효율적으로 학습하기 어렵다. 두 번째는 ‘Scrap’ 혹은 ‘Product’ 그룹과 같이 0값이 빈번하거나 동일한 값이 반복적으로 나타나는 변수들에 의하여 모델이 과적합 되거나 학습에 혼란을 겪을 수 있다. 이러한 문제를 해결하기 위해, 본 연구에서는 유사한 성격의 변수들을 기준으로 총 4개의 카테고리로나누었으며, 각 카테고리는 독립적인 FCN(fully connected network)을 통과하여 d 차원의 임베딩 벡터로 추출된다(<Figure 6>).

Table 2. Descriptions of the Variable Categories. The **Scrap** category consists of 12 variables, each representing the input weight of a specific scrap type. The **Product** category defines the process objectives, containing the target values for the final chemical components. The **Molten iron** category includes key features that quantify the initial thermal energy and properties of the hot metal. Finally, the **Oxygen** category comprises the primary control factors and energy inputs used to adjust final temperature.

Category name	Characteristic	Variables
Scrap	Input weight of scrap type	Scrap#1, Scrap#2 ... #12
Product	Target values for each chemical component	Target of C, Mn, P
Molten iron	Thermal state of molten iron	Initial temperature, iron weight, scrap weight, heat source, waiting time
Oxygen	Energy sources to reach final temperature	Total volume of oxygen, raw materials, slag state

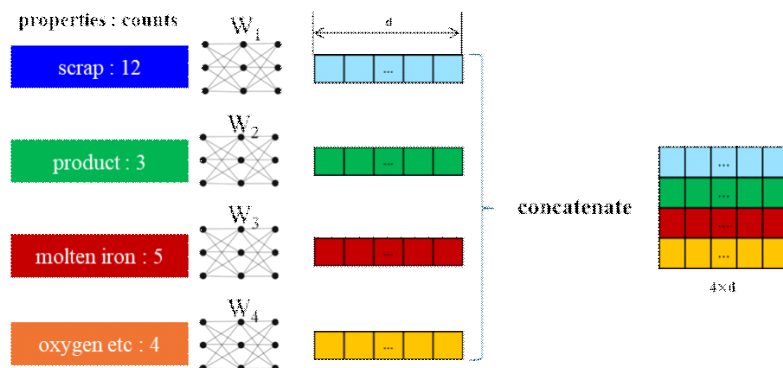


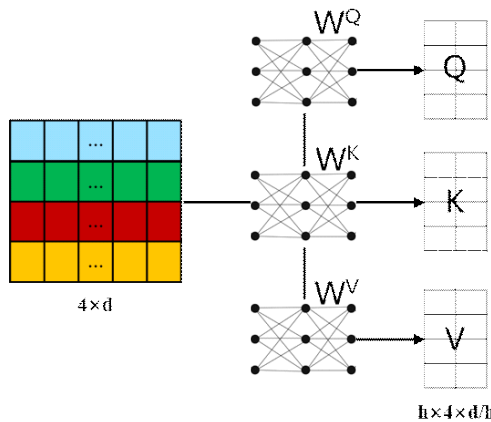
Figure 6. Schematic Illustration of the Feature Encoding Process: Independent Variables are Grouped Into Four Categories Based on Their Properties, Each Transformed Into a d-dimensional vector. The resulting vectors are concatenated to form a structured feature matrix for the attention mechanism

이 FCN 인코더는 비슷한 성질을 가진 고차원의 변수들을 의미론적으로 압축된 저차원의 밀도 있는 임베딩 벡터로 변환하는 역할을 수행한다. 결과적으로 모델은 개별 변수의 특성에 의존하는 대신 정제되고 통일된 벡터 표현을 입력으로 사용하게 되며, 이는 각 그룹이 최종 예측에 일관된 영향력을 행사하도록 유도하고, 기존 데이터의 노이즈로 인한 혼란 방지와 더불어 모델의 강건성을 높이기 위한 핵심적인 역할을 수행한다. 이러한 계층적 구조화의 타당성을 검증하기 위해 간단한 비교 실험을 수행하였다. 먼저 모든 입력 변수를 단일 벡터로 사용하여 예측하는 표준 선형 회귀 모델과, <Figure 6>로부터 얻은 (4×d)의 feature matrix를 Categorywise sum하여 얻은 벡터로 최종 예측을 수행하는 Category 회귀 모델의 성능을 <Table 3>에 비교하였다. 이 때 각 모델의 성능은 표준 제곱 오차(mean squared error, MSE)와 평균 절대 오차(mean absolute error, MAE)로 비교하였다.

Table 3. Regression performance comparison with and without variable categorization: The table confirms a significant performance improvement when the hierarchical category structure is applied, highlighted in bold.

Algorithms	MSE	MAE
Linear Regression	163.3	9.93
Linear Regression with Categories	144.1	9.37

<Table 3>의 결과에서 볼 수 있듯이, 계층적 카테고리 구조를 적용한 회귀 모델(Linear Regression with Categories)은 모든 변수를 단일 벡터로 사용한 표준 선형 회귀 모델(Linear Regression) 대비 MSE 기준 약 11.75%의 유의미한 성능 향상을 보였다. 이는 이질적이고 희소성이 높은 변수들을 계층적으로 그룹화하고 FCN 인코더로 압축하는 방식이, 모델의 예측 성능을 향상시키는 유효한 설계임을 보여주는 결과이다.



(2) Attention

비슷한 성질로 묶인 4개의 카테고리는 조업 조건에 따라 온도 예측에 미치는 기여도가 달라질 수 있으며, 특정 상황에서는 일부 카테고리가 예측에 있어 보다 중요한 역할을 수행하기도 한다. 이에 따라 본 연구에서는 각 카테고리 임베딩 간 상호 관계를 반영하고, 중요도를 동적으로 조정하기 위해 multi-head attention with category(MAC) 구조를 적용하였다. 먼저 각 카테고리로부터 생성된 임베딩 벡터들을 하나의 텐서 $H \in R^{B \times C \times d}$ 로 결합한다. 여기서 B 는 배치 크기, C 는 카테고리 수, d 는 임베딩 차원을 의미한다. 이후 이 텐서는 head별로 분할되어 각각의 선형 변환을 통해 query, key, value 텐서로 변환되며, <Figure 7>과 같이 다음의 attention 연산을 수행한다:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

여기서 d_k 는 head당 차원의 크기이며, softmax는 카테고리 축에 대해 적용되어, 각 카테고리 임베딩 간 상호 유사도를 기반으로 온도 예측에 중요한 카테고리가 강조되도록 attention weight를 학습한다. 이 때 attention 결과는 head별로 독립적으로 계산된 후, 이들을 병합(concatenation)하여 하나의 텐서로 구성한다. 이후 <Figure 8>과 같이 최종 출력을 다음과 같은 방식으로 계산한다:

$$z^t = Concat(head_1, \dots, head_h)W^O \quad (6)$$

여기에서 $head_i$ 는 식(5)에 의한 attention 계산 결과이고, h 는 head의 개수, W^O 는 연결된 head의 차원을 최종 임베딩 차원으로 조정하기 위한 선형 변환 행렬이다. 최종적으로, 각 시점 t 에서 생성된 attention 출력 z^t 는 <Figure 9>와 같이 카테고리 축(C)에 대해 합산하여 하나의 통합 벡터로 정리한다.

$$h_{attn}^t = \sum_{c=1}^C z_c^t \in R^d \quad (7)$$

$$softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V = \begin{matrix} \begin{matrix} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{matrix} \\ h \times 4 \times d \times h \end{matrix}$$

Figure 7. Multi-head attention with categories (MAC): the concatenated feature matrix is linearly projected to obtain query (Q), key (K), and value (V). Scaled dot-product attention captures dynamic relationships and relative importance among feature categories.

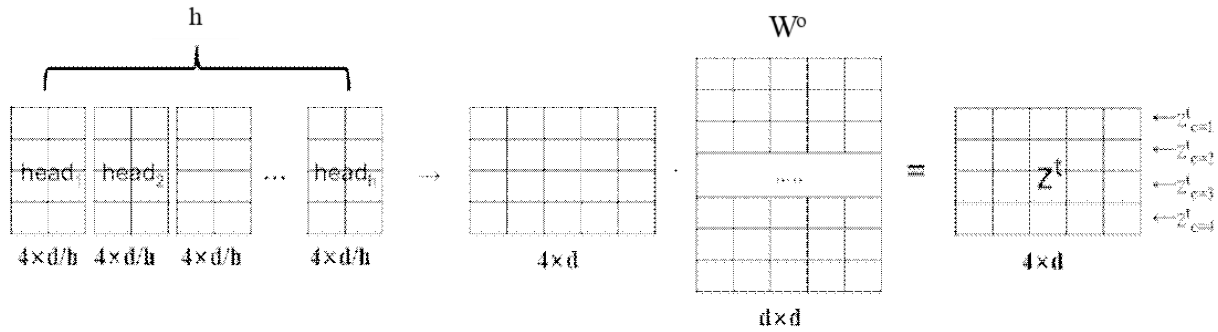


Figure 8. Final Output Projection of the Multi-head Attention Layer: Outputs from the Parallel Attention Heads are Concatenated and Transformed Through a Final Linear Transformation W^O , Producing the Output Vector z^t . Each row of z^t represents a context-aware embedding for one of the four feature categories

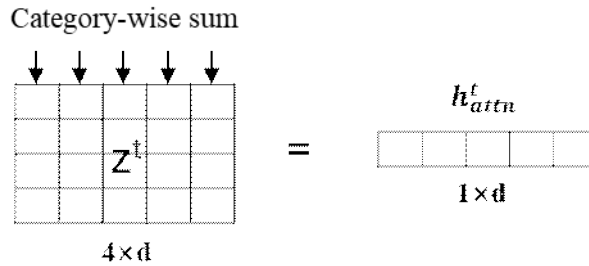


Figure 9. Category-wise Feature Aggregation: The Context-aware Embedding Matrix z^t , which Contains the Refined Information for All Four Feature Categories, is Aggregated Along its Category Axis to Produce a Fixed-size Context Vector Summarizing all Feature Information at the Current Time Step

여기서 z_c^t 는 시점 t 에서 카테고리 c 의 attention 연산을 거쳐 얻어진 벡터이며, h_{attn}^t 는 GRU cell에 입력되는 MAC구조의 최종 출력 벡터를 의미한다. 이러한 구조를 통해 모델은 조업 상황에 따라 각 카테고리의 중요도를 동적으로 판단하고, 온도 예측에 유효한 정보를 중점적으로 통합할 수 있다.

(3) Time Attention

절대 수명과 상대 수명은 각각 전로의 물리적 상태 변화에 따른 조업 특징의 변화와 이전 차지로부터 이어지는 잔류 슬

래그의 열원 정보를 반영해야 한다. 본 연구에서는 이를 위해 두 수명을 고정된 범주의 이산형 변수로 간주하고, <Figure 10>과 같이 lookup table 기반의 임베딩으로 표현하였다 (Mikolov *et al.*, 2013). 이를 통해 각 수명이 갖는 시간적 의미를 벡터 공간에 내재화하고, 이후 시계열 상에서의 중요도 계산 과정에 효과적으로 활용할 수 있도록 하였다. 이는 순환신경망의 hidden state 시퀀스 자체에 attention을 적용하는 표준적인 시계열 attention 방식에서 더 나아가, 전로 공정 자체의 독특한 물리적 시간을 이원화된 임베딩, 즉 전로 노후화 수명(절

[Absolute life embedding, ($4 \times d$)]			[Relative life embedding, ($50 \times d$)]	
Index	Vector, (dim= d ,)	Stage	Index	Vector, (dim= d ,)
0	[...]	Start	0	[...]
1	[...]	Early	1	[...]
2	[...]	Mid	...	[...]
3	[...]	End	49	[...]

Figure 10. Look-up Tables for Absolute and Relative Life Embeddings: Temporal Information is Encoded Using Two Embedding Matrices. Absolute life is mapped to one of four process stages and then to an embedding vector representing the converter’s long-term state. Relative life (the charge index within a cycle) retrieves a separate embedding vector capturing short-term sequential context.

대 수명)과 단기적인 조업 사이클 패턴(상대 수명)으로 직접 설계하고, 단순한 시간 순서가 아닌 전로 공정의 두 시간적 의미가 예측에 미치는 영향력을 모델이 직접 학습하도록 한다.

절대 수명과 상대 수명을 각각 표현하는 두 개의 벡터는 동일한 차원의 공간에서 합산되어 식(8)과 같이 하나의 시간 임베딩 $e_t \in R_t^d$ 을 형성한다:

$$e_t = e_t^{abs} + e_t^{rel} \quad (8)$$

여기서 e_t^{abs} 는 절대 수명 구간에 대한 임베딩이고, e_t^{rel} 는 해당 차지의 상대 수명에 대한 임베딩이다. 이렇게 생성된 시퀀스 $\{e_1, e_2, \dots, e_T\}$, ($T \leq 50$)는 시간 축을 따라 정렬된 시간 임베딩 시퀀스를 구성하며, 각 시점 t 의 중요도 점수는 다음과 같이 계산된다:

$$s_t = w^T e_t + b \quad (9)$$

여기서 $w \in R_t^d$ 는 학습 가능한 선형 변환 계수이며, $b \in R$ 는 bias 항이다. 이후 이 중요도 점수에 softmax 함수를 적용하여 전체 시퀀스에 대한 정규화된 가중치를 계산한다:

$$\alpha_t = \frac{\exp(s_t)}{\sum_{j=1}^T \exp(s_j)} \quad (10)$$

이렇게 얻어진 가중치 α_t 를 기반으로 전체 시퀀스에 대한 시간 문맥 벡터(time context vector) c 를 식(11)을 통해 계산하며, 이러한 일련의 과정을 <Figure 11>에 도식화 하였다.

$$c = \sum_{t=1}^T \alpha_t e_t \quad (11)$$

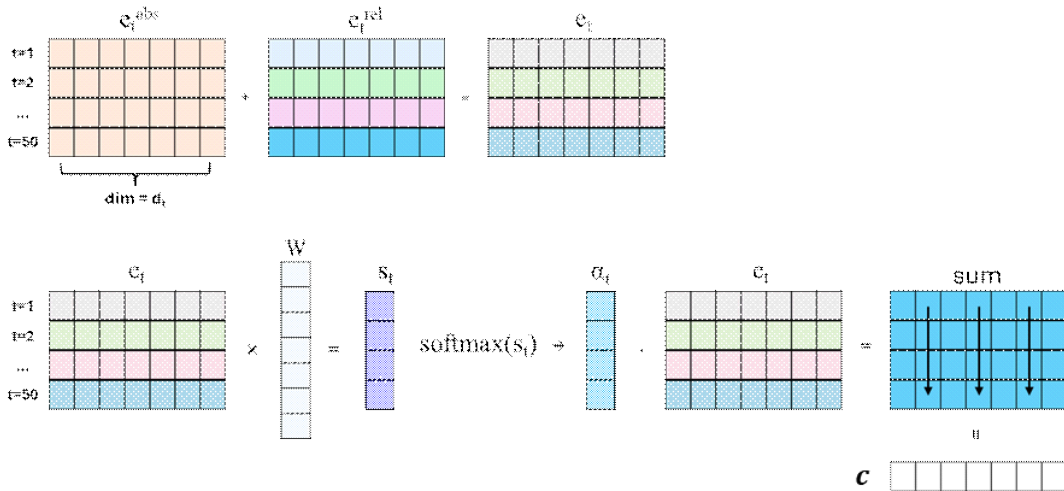


Figure 11. Computation of Time Context Vector c Via Time Attention: Absolute and Relative Life Embeddings are Combined Element-wise to form the Time Embedding Matrix e_t . An attention mechanism assigns weights α_t to each time step, and the final context vector c is obtained as a weighted sum, enabling the model to focus on the most relevant moments in the sequence.

한편, 추론 단계에서는 현재 시점 이후의 미래 정보가 모델에 영향을 미치지 않도록 하기 위해 마스킹 기법을 적용하였다. 유효 시퀀스 길이를 L 이라고 할 때, 시점 $t > L$ 에 대해서는 중요도 점수 s_t 를 강제로 음의 무한대로 설정하여 softmax 결과에서 해당 시점의 영향을 제거한다:

$$s_t = \begin{cases} w^T e_{t+b} & \text{if } t \leq L \\ -\infty & \text{if } t > L \end{cases} \quad (12)$$

따라서 모델은 오직 유효한 과거 시점($t \leq L$)까지의 정보만을 기반으로 time context를 구성하게 된다. 이 최종 time context는 앞서 3.2.2절에서 정의한 MAC 최종 벡터 h_{attn}^t 와 병합되어, 다음과 같이 최종 입력 벡터 h_{fused}^t 를 구성한다:

$$h_{fused}^t = \text{Concat}(h_{attn}^t, c) \quad (13)$$

이 병합과정은 <Figure 12>에 도식화 하였다. 이렇게 생성된 h_{fused}^t 는 시점 t 에서의 GRU cell 입력으로 전달되며, 이를 통해 시계열 및 조업 맥락 정보를 통합 반영한 온도 예측이 이루어진다.

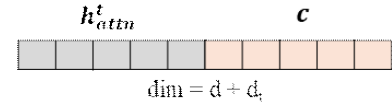


Figure 12. Fusion of MAC Output h_{attn}^t and Time Context Vector: The Aggregated Feature Vector from the MAC Module is Concatenated with the Time Context Vector from the Time Attention Module to form the Final Input to the Recurrent Layer

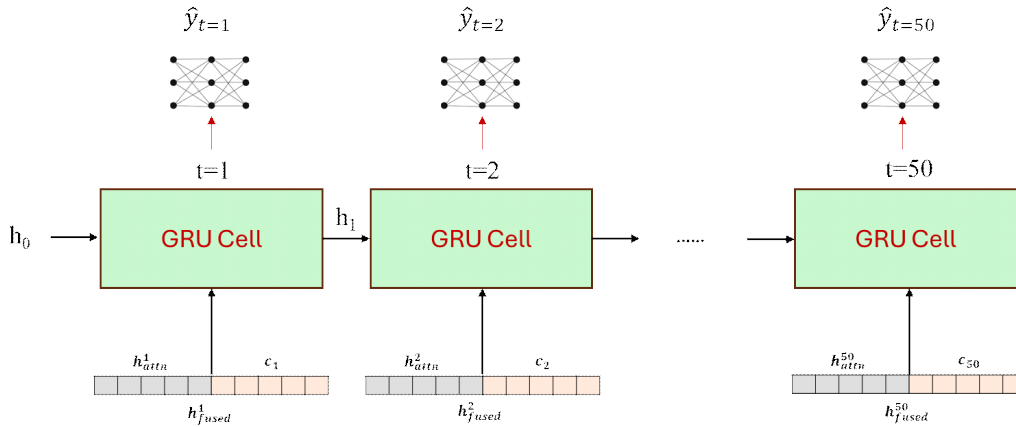


Figure 13. Recurrent Processing for Temperature Prediction at Each Time Step: The Fused Vector h_{fused}^t is the Input for the GRU Cell, which Updates its Hidden State h_t by Integrating the Previous Step $h_{(t-1)}$. This updated hidden state is fed into a final prediction layer to estimate the temperature \hat{y}_t .

(4) GRU 기반 시계열 예측 구조

본 연구에서는 전로 공정을 시간의 흐름을 갖는 연속적인 시스템으로 간주하고, 과거 차지의 조업 이력이 후속 차지의 온도 거동에 영향을 미친다는 점에 착안하여 시계열 예측 구조를 채택하였다. 시계열 정보를 효율적으로 학습하기 위해, GRU cell 기반의 순환 구조를 구성하였으며, 각 시점 t 의 입력으로는 MAC을 통해 생성된 조업 정보 요약 벡터 h_{attn}^t 와 time attention을 통해 얻어진 시간 정보 벡터 c 를 결합한 벡터 h_{fused}^t 를 사용하였다. 이 결합 벡터는 선형 변환층을 통과하여 GRU cell의 입력으로 전달되며, 시점 t 에서의 GRU 상태 갱신은 다음과 같이 이루어진다:

$$h_t = GRU(h_{fused}^t, h_{t-1}) \quad (14)$$

여기서 h_t 는 시점 t 에서의 은닉 상태이며, 이전 시점의 은닉 상태 h_{t-1} 과 함께 시계열 정보를 누적 반영한다. 이후, 매 t 시점의 중점 온도는 다음과 같이 GRU 출력 h_t 에 대한 선형 회귀를 통해 산출된다:

$$\hat{y}_t = W_y h_t + b \quad (15)$$

모델은 각 시점의 예측 온도 \hat{y}_t 와 실제 온도 y_t 간의 오차를 최소화하는 방향으로 학습되며, <Figure 13>과 같이 전체 시계열에 대해 예측 값을 출력한다.

4. 실험

4.1 실험 설계

본 연구의 목적은 제안한 시계열 기반 예측 모델이 실제 제

강 현장의 온도 예측 정확도를 개선할 수 있는지를 평가하는데 있으며, 이를 위해 다양한 전통적 회귀 기반 알고리즘들과의 비교 실험을 진행하였다. 실험은 특정 제강 공정 내 전로 조업 데이터를 기반으로 수행되었으며, 전체 약 35,000개의 조업 차지에 대해 3.1.2절에서 소개한 방법으로 사이클 단위의 시퀀스로 구성하였으며, 시퀀스 길이는 최대 50으로 제한하였다.

비교 대상은 조업 데이터를 개별 차지 단위의 독립 시행으로 간주하고 학습하는 전통적인 머신러닝 기반 회귀 알고리즘들이다. 가장 단순한 형태인 선형 회귀(linear regression, LR) 모델은 입력 변수의 선형 조합을 통해 목표값을 예측하며(Seber *et al.*, 2012), 이와 유사하게 ridge 및 lasso 회귀 모델은 각각 L2 및 L1 정규화를 통해 과적합을 억제하는 방식이다(Hoerl *et al.* 1970, Tibshirani 1996). ElasticNet은 L2 및 L1 정규화를 혼합하여 일반화 성능을 개선하고자 하는 알고리즘이다(Zou *et al.*, 2005). 결정 트리 기반 모델(decision tree, DT)은 입력 변수 공간을 기준에 따라 분할하며 예측값을 생성하고(Quinlan *et al.*, 1986), 이를 다수 조합하여 예측하는 랜덤 포레스트(random forest, RF)는 비선형 관계 모델링에 효과적인 앙상블 구조다(Breiman *et al.*, 2001). 이와 유사한 부스팅 계열 모델인 gradient boosting regressor(GBR)와 XGBoost는 잔차 기반의 반복적 성능 개선 방식으로 높은 예측력을 보이며(Friedman *et al.*, 2001), XGBoost는 그중에서도 정규화와 연산 최적화가 강화된 알고리즘이다(Chen *et al.*, 2016). 이외에도 거리 기반의 예측 방식인 k-NN regressor를 포함하여(Cover *et al.*, 1967), 다양한 유형의 회귀 알고리즘을 폭넓게 비교 대상으로 설정하였다. 이러한 알고리즘들은 모두 입력된 조업 데이터를 시계열로 해석하지 않고, 각 차지를 서로 독립적인 샘플로 간주하여 개별적으로 온도를 예측하도록 학습된다. 이는 실제 산업 현장에서 많이 채택되는 접근 방식이지만, 전로 조업의 연속성과 물리적 누적 특성을 반영하지 못한다는 한계가 있다. 따라서, 본 연구에서는 전로 조업의 시계열 연속성을 반영할 수 있

는 GRU 기반의 예측 모델을 제안하고, 이를 통해 시계열 예측 방법이 기존 독립 시행을 가정한 예측 방법 보다 더 정확하게 온도를 예측할 수 있음을 보이고자 한다.

4.2 평가 지표

예측 성능 평가는 실제 제강 현장에서 중요하게 고려되는 정량적 지표들을 중심으로 수행하였다. 기본적인 오차 측정 지표로는 평균제곱오차(MSE)와 평균절대오차(MAE)를 사용하였다. MSE는 예측값과 실제값의 차이를 제곱하여 평균한 값으로, 큰 오차에 민감하게 반응하며 다음과 같이 정의된다:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (15)$$

여기서 y_i 는 실제 온도, \hat{y}_i 는 예측 온도이며, N 은 전체 예측 시점의 개수이다. 식 (15)에서 보여주고 있는 MAE는 오차의 절대값을 평균한 값으로, 직관적인 해석이 가능하고 산업 현장에서 널리 활용되는 지표이다:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (17)$$

또한, 제강 공정에서는 온도 제어 능력을 정량적으로 평가하기 위하여 온도 적중률 지표를 운영하는데, 이 지표는 각 차지에 대해 목표 용강 온도에 얼마나 정확히 도달했는지를 나타내며, 다음과 같이 정의된다:

$$Accuracy_{\pm \delta} = \frac{1}{N} \sum_{i=1}^N \Pi (|y_i - \hat{y}_i| \leq \delta, \delta = 7 \text{ or } 10) \quad (18)$$

여기서 $\Pi(\cdot)$ 는 조건 만족 시 1, 그렇지 않으면 0을 반환하는 지시함수(indicator function)이다. 이를 통하여 예측 온도가 실제 온도의 $\pm 7^\circ\text{C}$ 또는 $\pm 10^\circ\text{C}$ 범위 내에 포함되는지를 정량적으로 평가할 수 있으며, 해당 모델이 목표 온도에 어느 정도로 근접하는지를 판단할 수 있다. 이 외에도, 모델의 실질적인 예측 성능을 보다 민감하게 파악하기 위해, 본 연구에서는 industrial MSE(iMSE)라는 별도의 오차 지표를 추가로 정의하였다. 이 지표는 예측 값이 실제 값과 $\pm 10^\circ\text{C}$ 이내일 경우 해당 예측을 정확하다고 간주하여 오차를 0으로 처리하고, $\pm 10^\circ\text{C}$ 이상 차이가 발생한 경우에만 MSE 방식으로 에러를 계산한다:

$$industrial\ MSE = \frac{1}{N} \sum_{i=1}^N \begin{cases} 0 & , \text{if } |y_i - \hat{y}_i| \leq 10 \\ (y_i - \hat{y}_i)^2 & , \text{otherwise} \end{cases} \quad (19)$$

이러한 방식은 온도 예측이 $\pm 10^\circ\text{C}$ 라는 현실적인 허용 오차 범위 내에 포함되는지를 기준으로 모델의 성능을 평가할 수

있도록 해준다. 기존 MSE는 예측값이 $\pm 10^\circ\text{C}$ 안에 들어왔더라도 오차가 계속 누적되기 때문에, 모델의 실질적인 “정답 적중 여부”를 판단하기에는 한계가 있다. 반면 iMSE는 정확히 $\pm 10^\circ\text{C}$ 를 기준으로 오차를 0 또는 유의미한 값으로 이분화하기 때문에, 온도 적중률에 따른 MSE 값의 변화를 보다 극적으로 반영하며, 모델의 예측 성능을 더욱 직관적이고 명확하게 파악할 수 있다. 이러한 특성은 특히 다양한 모델들 간의 미세한 성능 차이를 비교하는데 효과적인 판단 기준이 된다.

4.3 실험 결과

본 연구에서는 전로 조업의 시계열적 특성을 모델 구조에 반영한 순환 신경망 기반의 예측 모델을 제안하였다. 특히, 제안 모델은 단순한 순차 처리뿐만 아니라, 조업 정보 내 각 입력 변수군의 상대적 중요도를 학습하는 category attention과 조업의 시간적 흐름과 전로 수명에 따라 각 시점의 기여도를 반영하는 time attention 구조를 결합함으로써, 물리적인 공정 특성과 데이터 구조 양쪽을 모두 반영할 수 있도록 설계되었다. 이 모델은 MAC을 통해 입력 변수군 간의 중요도를 동적으로 조정하고, 동시에 time attention을 통해 조업 시점별 수명 정보를 반영함으로써 조업의 시간 축 상에서의 상대적 기여도를 학습한다. 이 때 하이퍼 파라미터는 grid search 방법으로 결정하였으며, <Table 4>와 같이 MAC의 hidden dimension과 head 수, 그리고 time attention의 dimension을 주요 조정 변수로 설정하였다. Batch size는 모든 실험에서 32로 고정하였다.

Table 4. Hyperparameter Search Space for the Proposed Model: Optimal Values were Selected Using Grid Search. The final configuration includes a MAC hidden dimension $d=128$, number of heads $h=32$, and time attention hidden dimension $dt=16$, highlighted in bold

Hyper parameter	values	Notation
MAC hidden dimension	[64, 128 , 256]	d
MAC number of head	[16, 32]	h
Time attention hidden dimension	[16 , 32]	d_t

모델의 성능을 검증하기 위해 비교 알고리즘을 세 그룹으로 나누어 <Table 5>에 통합하여 제시하였다. 첫 번째 그룹은 각 차지를 독립 시행으로 간주하고 학습하는 전통적인 머신러닝 회귀 알고리즘이다. 두 번째 그룹은 공정의 연속성을 단순 시계열로 간주하는 표준 순환 신경망이다. 마지막 세 번째는 본 연구에서 제안하는 dual attention 모델이다. 먼저 전통 회귀 모델 중 가장 우수한 성능을 보인 XGB regression과 표준 시계열 모델인 GRU 및 long short term memory(LSTM, Hochreiter *et al.*, 1997)를 비교한 결과, GRU와 LSTM이 XGB regression 대비 모든 지표에서 더 우수한 성능을 보였다. 이는 전로 조업을

Table 5. Performance comparison between the proposed model, recurrent neural network models and baseline regression algorithms: the proposed MAC-time attention model outperforms all the other methods across all evaluation metrics, with the best results highlighted in bold

Algorithms		MSE	iMSE	MAE	R2	$\pm 10^\circ\text{C}(\%)$	$\pm 7^\circ\text{C}(\%)$
Machine Learning	Linear Regression	163.3	144.8	9.93	0.381	59.4	43.4
	Ridge	163.3	144.9	9.93	0.381	59.4	43.4
	Lasso	193.7	176.9	10.94	0.265	53.8	38.7
	Elastic Net	170.1	152.7	10.20	0.355	57.1	42.4
	Decision Tree	171.7	154.1	10.28	0.349	57.3	42.5
	K-NN regression	182.3	164.9	10.57	0.309	56.0	41.3
	RF regression	150.5	132.2	9.57	0.429	60.9	45.4
	GBR	149.3	126.1	9.53	0.434	60.8	45.5
	XGB regression	144.3	131.3	9.35	0.453	61.7	46.4
Recurrent Neural Network	LSTM	140.8	122.5	9.33	0.462	61.8	46.7
	GRU	140.5	121.8	9.31	0.463	61.9	46.8
Proposed	MAC-time attention	130.7	110.9	8.96	0.500	64.4	47.4

독립적인 사건이 아닌 연속적인 시계열로 가정하는 것만으로도 예측 성능이 유의미하게 향상됨을 입증하며, 본 연구의 기본 가설이 타당함을 뒷받침한다.

다음으로, 표준 순환 신경망 모델의 성능을 제안 모델과 비교하였을 때, 제안 모델이 모든 성능 지표에서 전통 회귀 모델 대비 뛰어난 성능을 나타내었으며, 특히 $\pm 10^\circ\text{C}$ 및 $\pm 7^\circ\text{C}$ 온도 적중률에서 높은 정확도를 기록하였다. 이는 조업 진행에 따라 누적되는 전로 공정의 특성과 연속되는 차이 사이에 남아 있는 열원의 비선형적 영향을 모델이 효과적으로 학습하고 반영했음을 보여주는 결과라 할 수 있으며, 복잡한 제강 공정에서 종점 온도를 예측할 때, 모든 시점과 변수의 정보를 동일한 가중치로 처리하는 단순 순환 신경망 구조만으로는 전로의 시계열 특징을 반영하는 데 명백한 한계가 있음을 시사한다. 반면, 본 연구에서 제안하는 dual attention 기반 순환 신경망은 예측에 결정적인 영향을 미치는 특정 변수 그룹이나 특정 시점의 정보에 동적으로 집중하여 가중치를 부여함으로써, 단순 순환 신경망으로는 할 수 없었던 복잡한 조업 패턴을 효과적으로 포착할 수 있다는 것을 확인할 수 있다.

추가적으로, 각 attention 모듈의 효과를 검증하기 위해, time attention 구성 요소를 제거한 MAC-GRU 구조와, category attention을 제거한 Time only GRU 시계열 순환 신경망 모델을 별도로 비교하였다. MAC GRU 모델은 카테고리 임베딩 간의 상호 작용만을 고려하고, 시간 축 상의 전로 수명 정보를 활용하지 않는 구조로, GRU 입력에 time context 벡터를 포함하지 않는 구조이다. 즉, time attention 구조가 제거된 경우, 절대 수명 및 상대 수명은 단순히 1씩 증가하는 숫자의 나열로만 입력되며, 이에 대한 임베딩 표현은 존재하지 않는다. 이로 인해 두 수명 정보는 모델 내에서 실질적인 의미를 갖기 어려우며, 결과적으로 GRU 셀이 시점 증가에 따라 생성하는 hidden state만으로 전로의 열 축적이나 원료 투입 패턴과 같은 복잡한 조업 특성을

모두 내재적으로 학습해야 하는 구조가 된다. 이러한 구조적 제약으로 인해 <Table 6>의 결과를 보면 MAC GRU는 MAC-time GRU 대비 모든 지표에서 성능이 다소 저하됨을 확인할 수 있다. 이는 시간 축 상의 각 조업 시점이 온도 예측에 미치는 기여도를 모델이 학습하도록 설계된 time attention 구조가 단순한 시계열 순서 처리만으로는 포착하기 어려운 전로의 절대 수명 및 상대 수명에 대한 시간 맥락 정보를 보완하고 있다는 것을 보여준다. 이는 time attention이 수명 정보를 시점별로 적절히 강조함으로써 전로 공정의 시계열적 및 누적 특성을 효과적으로 반영하는 데 중요한 역할을 수행함을 보여준다.

Time Only GRU 모델은, time attention 모듈만을 유지한 채로, 변수들을 그룹핑하여 attention을 수행하는 MAC 구조를 제외한 것으로, 모든 변수를 일반적인 회귀 모델과 같이 단일 벡터로 처리하는 모델이다. 즉, category attention이 제거된 경우, 비슷한 성질의 변수들이 용강의 온도에 동일한 영향력을 행사한다는 내재적 관계를 모델이 명시적으로 활용하지 못하고, 희소성이 높은 변수들과 노이즈로 인해 학습에 혼란을 겪을 위험을 안고 있다. 실험 결과, Time only GRU 모델은 MAC-time GRU 대비 모든 지표에서 성능이 저하됨을 확인할 수 있다. 이는 MAC 구조가 3.2.1절에서 제기한 입력 변수의 이질성과 희소성 문제를 효과적으로 해결하고, 정제되고 압축된 특징 벡터를 생성함으로써 모델의 안정적인 학습을 돕는 매우 중요한 역할을 수행하고 있음을 보여준다.

마지막으로, 절대 수명과 상대 수명이 모델에 기여하는 정도를 개별적으로 파악하기 위해, 각 수명에 대한 임베딩 벡터를 분석하였다. <Figure 14>는 절대 수명 임베딩 벡터의 각 단계에 대한 L2-norm을 시각화 한 것으로, 시작 단계의 임베딩 벡터 크기가 가장 크고, 이후 초기, 중기 구간에서 점차 감소한 뒤, 말기 구간에서는 일정 수준에서 유지됨을 확인할 수 있다.

Table 6. Ablation Study on the Attention Modules: Incorporating Dual Attention (MAC-time GRU) Improves Performance Across all Metrics Compared to the Model Without Time Attention (MAC GRU) and without Category Attention Modules (Time only GRU)

Algorithms	d, h, d_t	MSE	iMSE	MAE	R2	$\pm 10^\circ\text{C}(\%)$	$\pm 7^\circ\text{C}(\%)$
MAC-time GRU	128, 32, 16	130.7	110.9	8.96	0.500	64.4	47.4
MAC GRU	128, 32	136.1	117.6	9.15	0.480	62.5	47.0
Time only GRU	128, -, 16 -	138.7	120.6	9.32	0.461	61.2	46.0

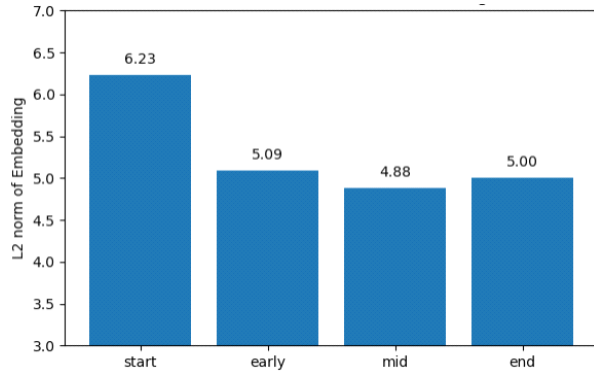


Figure 14. L2-norms of Absolute Life Embedding Vectors: The L2-norm Peaks at the ‘Start’ Stage, Reflecting the Distinct Thermal and Chemical Conditions of Early Operation. This indicates the model effectively captures domain-specific characteristics.

이는 실제 조업에서 전로 가동 시작 단계에서 전로 내부에 축조된 내화 벽돌의 열축적이 충분하지 않고 내용적이 작기 때문에, 산소 취입을 위한 설비 제어 방식, 원료 투입 방법 등을 다르게 설정하는 경향과 일치하며, 모델이 이러한 조업 기조의 차이를 효과적으로 학습했음을 보여준다.

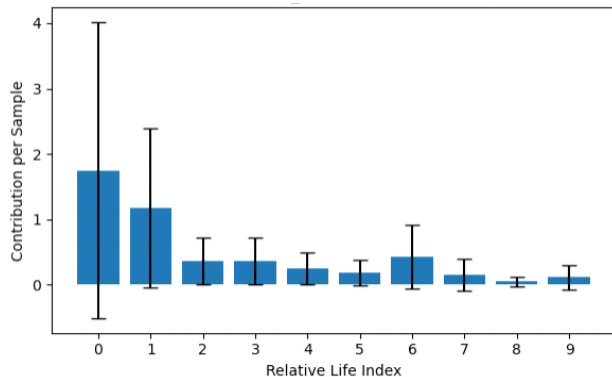


Figure 15. L2-norms of Relative Life Embeddings for the First 10 Charges: Higher Norms at the Beginning Indicate that the Model Captures the Greater Complexity and Importance of Initial Charges in a New Cycle

<Figure 15>는 연속 조업 사이클 내 초반 차지(10번째)의 상

대 수명 임베딩 벡터의 L2-norm을 시각화 한 결과로, 절대수명과 마찬가지로 첫 번째 순서의 임베딩 벡터의 크기가 가장 크고 상대 수명이 증가할수록 점점 감소하였다. 상대 수명 첫 번째 차지는 실제 조업에서 잔류 슬래그를 모두 버린 후 대기 상태에 있다가 다시 조업을 재개한 경우로, 전로 내부에 축적되었던 열원이 복사열로 일부 소실되고, 잔류 슬래그의 대체를 위해 원료를 추가로 투입해야 하므로 전체적으로 열적 손실이 우세한 조업 조건이다. 이후 잔류 슬래그 양을 일정한 수준으로 유지하면서 조업의 편차를 줄이는데, 이와 같은 조업 기조가 상대 수명 임베딩에 효과적으로 내재화되었음을 확인할 수 있었다.

5. 결론

본 연구는 전로 제강 공정에서의 최종 용강 온도 예측 정확도를 향상시키기 위해, 조업의 연속성과 시간 축 상의 맥락 정보를 반영한 시계열 기반 예측 모델을 제안하였다. 기존의 접근 방식은 각 차지를 독립된 작업 단위로 간주하고 예측을 수행했기 때문에, 조업 간 잔류 슬래그 및 전로 축열 상태와 같은 연속 조업 특성이 온도 거동에 미치는 영향을 반영하지 못하는 한계가 있었다. 이에 본 연구는 조업 흐름을 시계열로 재구성하고, 조업 조건을 기능적으로 구분한 category attention, 그리고 전로의 절대 수명 및 상대 수명 정보를 내재화한 time attention을 통해, 시점별 조업 맥락을 반영할 수 있도록 설계하였다. 실험 결과, 제안한 모델은 전통적인 회귀 기반 모델 대비 모든 평가지표에서 우수한 성능을 보였으며, 특히 $\pm 10^\circ\text{C}$ 및 $\pm 7^\circ\text{C}$ 내의 온도 적중률에서도 유의미한 향상을 보였다. 아울러 추가 실험 및 임베딩 분석을 통해, 전로의 사용 수명 및 조업 흐름상의 위치 정보가 온도 거동에 중요한 영향을 미치며, 본 모델이 이를 효과적으로 학습하고 활용함을 확인하였다. 이러한 연구 결과는 단순한 예측 정확도 향상을 넘어, 향후 조업 자동화 및 제어 최적화에 기여할 수 있는 기반 기술로 확장 가능하다. 특히, 목표 온도 달성을 위한 원료 투입량 및 산소 취입 조건의 사전 제어, 불필요한 열원 투입 방지, 공정 변동성 완화 등에도 효과적으로 활용될 수 있을 것으로 기대된다. 본 연구의 온도 예측 모델을 기반으로, 향후에는 목표 온도가 주어졌을 때 필요한 산소, 원료의 양을 실시간으로 판단하는 공정 자동화 시스템 구축을 위한 후속 연구를 진행할 예정이다. 이를

통해 단순한 예측을 넘어 실질적인 공정 제어 최적화까지 이어질 수 있는 지능형 전로 자동화 시스템으로 발전시킬 수 있을 것으로 기대된다.

참고문헌

- Breiman, L. (2001), Random forests, *Machine Learning*, **45**(1), 5-32.
- Chen, T. and Guestrin, C. (2016), XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014), Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, arXiv preprint arXiv:1412.3555.
- Cover, T. and Hart, P. (1967), Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, **13**(1), 21-27.
- Feng, K., Yang, L., Su, B., Feng, W., and Wang, L. (2022), An Integration Model for Converter Molten Steel End Temperature Prediction Based on Bayesian Formula, *Steel Research Int.*, **93**, 2100433.
- Friedman, J. H. (2001), Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, **29**(5), 1189-1232.
- Gu, M., Xu, A., Yuan, F., He, X., and Cui, Z., (2021), An Improved CBR Model Using Time-series Data for Predicting the Endpoint of a Converter, *ISIJ International*, **61**(10), 2564-2570.
- Hochreiter, S. and Schmidhuber, J. (1997), Long short-term memory, *Neural Computation*, **9**(8), 1735-1780.
- Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12**(1), 55-67.
- Jo, H. J., Hwang, H., Phan, D., Lee, Y., and Jang, H. (2019), Endpoint Temperature Prediction model for LD Converters Using Machine-Learning Techniques, *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, Tokyo, Japan, 2019, 22-26.
- Li, X., Liu, J., Zhao, Y., and Wang, H. (2023), Carbon Neutrality Orientates the Reform of the Steel Industry, *Journal of Iron and Steel Research International*, **30**(7), 1543-1553.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv:1301.3781.
- Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., and Cottrell, G. (2017), A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction, *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2627-2633.
- Qiu, X., Zhang, R., and Yang, J. (2024), Prediction of BOF endpoint carbon content and temperature via CSSA-BP neural network model, *J.Iron Steel Res. Int.*, **32**, 578-593
- Quinlan, J. R. (1986), Induction of decision trees, *Machine Learning*, **1**(1), 81-106.
- Seber, G. A. F. and Lee, A. J. (2012), *Linear regression analysis*, John Wiley & Sons.
- Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1), 267-288.
- Zou, H. and Hastie, T. (2005), Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(2), 301-320.

저자소개

한경석 : 한양대학교 신소재공학과에서 2013년 학사 학위를 취득, 2013년 Posco에 입사 후 근무하다가 2024년부터 고려대학교 산업경영공학과 석사과정에 재학 중이다. 연구 분야는 Timeseries Regression, Deep Learning이다.

김성범 : 고려대학교 산업경영공학부 교수로 2009년부터 재직하고 있으며, 인공지능공학연구소 소장, 기업산학협력센터 센터장, 한국데이터마이닝학회 회장을 역임했다. 미국 University of Texas at Arlington 산업공학과에서 교수를 역임하였으며, 한양대학교 산업공학과에서 학사학위를 미국 Georgia Institute of Technology에서 산업시스템공학 석사 및 박사학위를 취득하였다. 인공지능, 머신러닝, 최적화 방법론을 개발하고 이를 다양한 공학, 자연과학, 사회과학 분야에 응용하는 연구를 수행하고 있다.