AISTech 2024 — Proceedings of the Iron & Steel Technology Conference
6–9 May 2024, Columbus, Ohio., USA
DOI: 10.33313/388/225

Generative AI for Process Modeling in the Steel Industry

R. Marquez, E. Morales, A. Alvarez

ECON Tech
Corporativo Angelópolis, Paseo Opera 2-Office 206, Lomas de Angelópolis, 72830 Puebla, Pue., Mexico
Phone: +52 222 478 8282
Email: richard.marquez@econ-tech.com

ABSTRACT

This paper investigates the generation of realistic process data by utilizing adversarial networks to emulate actual steel process information. The objective is to create data resembling process data without discernible differences. Obtaining accurate process models for steel production is complex due to multiple challenges. Nevertheless, such data can prove invaluable for training operations personnel, simulating scenarios, testing software systems, and exploring novel control schemes. This research explores various generative modeling approaches to gain a deeper understanding of the constraints and practical implications of these models in steel environments, focusing on the application to a public dataset on continuous casting machine.

Keywords: Generative Adversarial Networks (GANs), Process Modeling, Steel Industry Simulation, Data Generation Techniques

INTRODUCTION

Artificial Intelligence (AI) is pervading almost any industry. The steel industry has been at the forefront of adopting new technologies to improve efficiency, reduce costs, and enhance product quality. In the last decades, there has been a growing interest in using AI to optimize steel production processes. Accurate process modeling for steel production is complex due to various challenges, including the complexity of the processes involved, variability in data, and the need for real-time monitoring and control. From those, data remains at the core of Industry 4.0 and AI in the steel industry.

Raw data from the automation layer requires some preprocessing (manipulation and curation) to be usable.[1] In particular, we sometimes need more data for process modeling or to train our AI models effectively, but a lack of data is present (missing values or missing samples). That's when we can leverage statistical data augmentation (dataset expansion) mechanisms to generate more data and increase diversity. Sagasti (2018), for example, considers eliminating data samples or features and simple imputation.[1] Data augmentation techniques involve creating new data samples from the existing ones through various methods like geometric transformations, noise injection, and mixing samples. On the other hand, a closely related concept, synthetic data generation, involves creating entirely new data samples not present in the original dataset. This approach consists of learning the underlying data distribution and sampling new synthetic data points from it.

Data augmentation is a powerful technique used to increase the size and diversity of datasets, which can be particularly useful in machine learning tasks involving steel processes. It can be used to improve the performance of machine learning models, especially in cases where a class imbalance exists; class imbalance creates bias (bad generalization), which is the main issue from the point of view of AI classification algorithms. However, managers and process engineers can employ this technique for simulation and decision-making purposes: assisting a digital twin, testing software systems (and even new control schemes), or training personnel on (faulty) simulated environments of the steel manufacturing operation; this can be particularly beneficial in facilitating decisions for less experienced operators, reducing errors, and enhancing operation accuracy, which ultimately leads to cost savings and CO_2 emission reductions. Furthermore, synthetic data can help a steel company share its behavior to the public without revealing confidential industrial data, thus protecting the privacy of sensitive information.

There are several approaches to data augmentation:

- Distribution-based augmentation generates synthetic data samples based on the input data features' known or assumed probability distributions. For steel processes, this could include modeling the distributions of various process

