Perlin Noise and Cosine Scheduler in Diffusion Models for Non-normal Data Imputation

Ari Wibisono, author

Deskripsi Lengkap: https://lib.ui.ac.id/detail?id=9999920567801&lokasi=lokal

----

Abstrak

Imputation is a critical method for enhancing dataset quality, essential for ensuring accurate analysis and insights. This research proposes an advanced imputation algorithm utilizing a Diffusion Model enhanced with Perlin noise generation. Perlin noise is introduced at each step of the diffusion process, paired with a cosine scheduler to optimize performance. Its smooth and continuous nature makes it well-suited for non-normal data by introducing gradual variations that align closely with the underlying structure of the data. Our approach demonstrates improvements in imputing non-normal data, validated through tests on ten real datasets and comparisons to imputation methods based on diffusion models. The results highlight a marked reduction in RMSE error values, achieving up to a 10% improvement, showcasing the efficacy of Perlin noise and the scheduler. However, the computational cost of Perlin noise, with its exponential complexity is significantly higher than Gaussian noise, making it particularly impactful in high-dimensional datasets. Despite this, the superior accuracy achieved justifies the cost for complex, non-normal datasets, offering a robust alternative to standard imputation methods. These dissertation findings providing insights into the usage of Perlin noise-based imputation for real-world applications.
......Imputation is a critical method for enhancing dataset quality, essential for ensuring accurate analysis and insights. This research proposes an advanced imputation algorithm utilizing a Diffusion Model enhanced with Perlin noise generation. Perlin noise is introduced at each step of the diffusion process, paired with a cosine scheduler to optimize performance. Its smooth and continuous nature makes it well-suited for non-normal data by introducing gradual variations that align closely with the underlying structure of the data. Our approach demonstrates improvements in imputing non-normal data, validated through tests on ten real datasets and comparisons to imputation methods based on diffusion models. The results highlight a marked reduction in RMSE error values, achieving up to a 10% improvement, showcasing the efficacy of Perlin noise and the scheduler. However, the computational cost of Perlin noise, with its exponential complexity is significantly higher than Gaussian noise, making it particularly impactful in high-dimensional datasets. Despite this, the superior accuracy achieved justifies the cost for complex, non-normal datasets, offering a robust alternative to standard imputation methods. These dissertation findings providing insights into the usage of Perlin noise-based imputation for real-world applications.