

Instruction-based Time Series Editing

Keywords: time series editing, text-time-series fusion, multi-modality, contrastive learning.

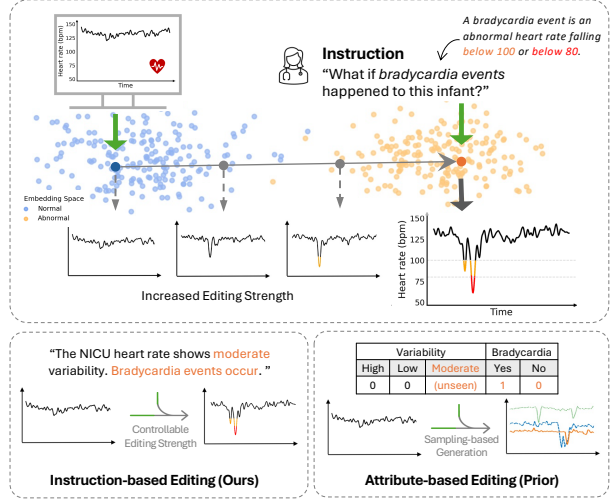
Time series editing (TSE) is a growing area that enables fine-grained modification of time series under specific conditions while preserving original patterns. This task is especially important in domains like healthcare, where practitioners may need to explore what-if scenarios, such as how a patient's vital signs would evolve under adverse events. Existing diffusion-based TSE methods rely on hardcoded attribute vectors and produce rigid all-or-nothing edits, limiting their flexibility and interpretability [1, 2]. These constraints prevent nuanced, semantically rich interactions and prohibit fine control over edit strength.

To address these limitations, we introduce **Instruction-Based Time Series Editing**, a new task where users guide edits via natural language instructions. Rather than using predefined attributes, users specify desired changes directly in text, enabling a broader and more flexible set of editing operations. This shift reflects real-world settings, where time series are often accompanied by unstructured notes or event descriptions. Instructional edits also allow gradual and controllable application of changes, supporting hypothesis generation and exploratory analysis.

We propose **InstructTime**, the first model for instruction-based time series editing. InstructTime maps both time series and natural language instructions into a shared multimodal embedding space on a unit-length hypersphere using contrastive learning. It then decodes interpolated embeddings to generate edited time series with varying degrees of instruction influence. This design allows InstructTime to support *controllable editing strength*, *compositional multi-condition instructions*, and *generalization to unseen instructions* including via few-shot adaptation.

Our architecture includes multi-resolution time series encoders to capture both global trends and local patterns, addressing the fact that different instruction components may relate to different temporal resolutions. We show that InstructTime performs high-fidelity edits across both synthetic and real-world datasets. Compared to state-of-the-art diffusion-based methods, it achieves stronger control on editing strength, better semantic alignment with instructions, and competitive or superior quantitative performance. We also evaluate a non-instruction version of InstructTime using categorical attributes to ensure a fair comparison with existing approaches and find that it remains highly competitive.

Overall, InstructTime bridges the gap between time series editing and natural language, providing a more expressive, interpretable, and flexible framework for controlled time series generation. Our contributions include: (1) we introduce a new TSE paradigm using free-text instructions; (2) we propose a novel multimodal model for controllable edits; and (3) we demonstrate the effectiveness of our method on diverse benchmarks. This work is currently under review at KDD 2026.



[1] Jing, Baoyu, et al. "Towards editing time series." *Advances in Neural Information Processing Systems* 37 (2024): 37561-37593.

[2] Narasimhan, Sai Shankar, et al. "Time weaver: A conditional time series generation model." *arXiv preprint arXiv:2403.02682* (2024).