Jesse Waite

jdwaite@wsu.edu

(425) 877-5936

Dear Information Systems editor or editors,

I wish to submit the included research article “Anomaly Detection in Processes Represented as a Graph” for publication. This work is a shorter version of my computer science master’s thesis at Washington State University, and was co-authored by my adviser, professor Lawrence Holder, PhD. This work has not been published elsewhere and was self-funded.

This work represents a new application of the SUBDUE graph compression method for process mining data. It provides a demonstration of the effectiveness of these methods via anomaly detection experiments using several controlled datasets and one real-world dataset, as well as a comparison with an existing algorithm. This work was difficult to produce, given both the lack of process mining data with controlled characteristics, and the complexity of constructing algorithms in the high-dimensional space of graphs that generalize to different graphical data distributions. A large part of its production was simply developing the experiments to adequately characterize and encompass these distributions, which is often omitted from similar works.

Recent works have confirmed the utility of graph-compression methods for qualitative analyses of processes, but none have incorporated these into a quantitative method for process analysis with evaluation over a range of process-oriented data distributions. Our work advances a trend in process mining toward machine learning approaches, which cultivate the view of process data as a distribution over graphical structure. The attached research paper advances this perspective by applying a rigorous algorithmic approach, but also lends to this broader trend by formalizing that view. The contributions are twofold since the method described in this work proved successful on synthetic and real-world data under a range of conditions, but it can also guide more advanced machine learning approaches that leverage the substructure-distribution view to inform their architecture. Such approaches might ultimately lend a blackbox approach to process analysis and anomaly detection without the manual engineering inherent in our work.

Thus, our work demonstrates the utility of graph compression for process analysis and anomaly detection through the development of an anomaly detection method with strong performance on several datasets. Additionally, the work advances the view of process data as a distribution of substructures, providing a valuable characterization for practitioners wishing to improve on our manually-engineered approach with more advanced methods.

Thank you for your time and consideration, and please feel free to send me any inquiries.

Sincerely,

Jesse Waite, M.S.