

very poor fitness and 1 for perfect fitness. Our result showed that the average fitness/case increased and standard deviation also increased to show an increase in the variation of cases in the log.

6. Conclusion and Future Work

In this paper, we applied data aware-process mining as a process-data centric approach to improve data quality in the early life cycle stages of big data applications by remodeling the process to capture data more accurately. Data-aware process mining is an approach presented by M. de Leoni et al. in [14]. We applied their approach and improved it by adding compliance rules for conformance checking.

The proposed method has been implemented using existing plug-ins in ProM and evaluated on a real-life event log obtained from the Dutch Financial Institute. The evaluation of our method for early detection of errors in process flow using Petri nets with data was successful in providing new and valuable insights in detecting introductory points of error. However, our filtering method also failed in other areas of performance as indicated by calculation time statistics, and synchronous event classes' statistics.

In future work, this team would like to implement the suggested diagnostic data-aware detection algorithm as a Prom plug-in which is an integrated algorithm for checking the alignment of a Petri Net with data and finally generating a verification report for the workflow designer.

References

- [1] Alexandros Labrinidis, H.V Jagadish, "Challenges and Opportunities with Big Data", Proceedings of the VLDB Endowment, Istanbul, Turkey, 2012.
- [2] Eckerson W, "Data quality and the bottom line: Achieving business success through a commitment to high quality data", The Data Warehousing Institute, 2002.
- [3] Barna Saha, Divesh Srivastava, "Data Quality: The other Face of Big Data", ICDE Conference, 2014.
- [4] Mark Madsen, "The Challenges of Big Data & Approaches to Data Quality", Third Nature Inc, 2013.
- [5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity", McKinsey Global Institute, 2011.
- [6] Michael Schroeck, et al, "Analytics: The real world use of big data", 2012. Available: http://www.ibm.com/systems/hu/resources/the_real_world_use_of_big_data.pdf [Accessed: January 21, 2015].
- [7] S.W. Sadiq, M.E. Orlowska, W.Sadiq, and C. Foulger, "Data Flow and Validation in Workflow Modelling", In Fifteenth Australasian Database Conference (ADC), Dunedin, New Zealand, volume 27 of CRPIT, pp 207–214, Australian Computer Society, 2004.
- [8] S.X. Sun, J.L. Zhao, J.F. Nunamaker, and O.R. Liu Sheng, "Formulating the Data Flow Perspective for Business Process Management", Information Systems Research, 17(4):374–391, 2006.
- [9] S.W. Sadiq, M.E. Orlowska, W.Sadiq, and C. Foulger, "Data Flow and Validation in Workflow Modelling", In Fifteenth Australasian Database Conference (ADC), Dunedin, New Zealand, volume 27 of CRPIT, pp 207–214, Australian Computer Society, 2004.
- [10] S. Fan, W.C. Dou, and J. Chen, "Dual Workflow Nets: Mixed Control/Data-Flow Representation for Workflow Modeling and Verification", In Advances in Web and Network Technologies, and Information Management (APWeb/WAIM 2007 Workshops), volume 4537 of Lecture Notes in Computer Science, pp 433–444, Springer-Verlag, Berlin, 2007.
- [11] W.M.P. van der Aalst, "The Application of Petri Nets to Workflow Management", The Journal of Circuits, Systems and Computers, 8(1):21–66, 1998.
- [12] M.H. Sundari, A.K. Sen, and A. Bagchi, "Detecting Data Flow Errors in Work-flows: A Systematic Graph Traversal Approach", In 17th Workshop on Information Technology & Systems (WITS-2007), Montreal, 2007.
- [13] Nikola Trocka, Wil M.P. van der Aalst, and Natalia Sidorova, "Data-Flow Anti-Patterns: Discovering Dataflow Errors in Workflows", LNCS 5565, 2009.
- [14] Massimiliano de Leoni and Wil M.P. van der Aalst, "Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments", ACM 978-1-4503-1656-9, 2013.
- [15] Elham Ramezani, Vladimir Gromov, Dirk Fahland, and Wil M. P. van der Aalst, "Compliance Checking of Data-Aware and Resource-Aware Compliance Requirements", On the Move to Meaningful Internet Systems: OTM 2014 Conferences Lecture Notes in Computer Science Volume 8841, pp 237-257, 2014.
- [16] Wil van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.F, "Replaying history on process models for conformance checking and performance analysis", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(2), pp 182–192, 2012.
- [17] M. de Leoni, W. M. P. van der Aalst, B. F. Van Dongen, "Data- and Resource-Aware Conformance Checking of Business Processes", In 15th International Conference on Business Information Systems, volume 117 of LNBIP, pp 48-59, Springer Verlag, 2012.