Process mining and deep learning in the natural sciences and process development

PI: Janos Abonyi , www.abonyilab.com

NKFIH-OTKA 116674; 02/2016-01/2020

The aim of the project was to work out deep learning, time series and sequence analysis algorithms to support the development and operation of process technologies.

Deep learning and process/sequence mining for pocess monitoring and alarm management

https://www.abonyilab.com/data-science/deep-learning

https://www.abonyilab.com/systems-engineering/alarm-management

Following the work plan, we worked out sequence mining and deep learning algorithms to solve problems in the field of chemical process engineering.

- We developed a Python - KERAS - TENSORFLOW based solutions (https://github.com/abonyilab/understanding_RNN)
- We extended the benchmark simulator of a vinyl acetate production technology to generate easily reproducible results and stimulate the development of our deep learning and process mining algorithms (focused on development of alarm management solutions) (https://github.com/abonyilab/VACsimulator)
- We processed the historical alarm data of the delayed cracking unit of MOL Ltd.
- We designed a sequence to sequence based process moonitoring and frequent pattern mining algorithms to help the analysis of complex chemical technologies and applied at MOL.

Scientific results:

Deep learning-based algorithms are suitable for root-cause analyis and visualisation of alarm signals. We developed a method to visualize deep neural networks. The developed algorithm can help the extraction of hidden relationships between the variables of large process datasets. We presented that the visualization of deep learning models can be used for root-cause analysis

- Gyula Dorgo,Peter Pigler,Janos Abonyi: Understanding the importance of process alarms based on the analysis of deep recurrent neural networks trained for fault isolation, JOURNAL OF CHEMOMETRICS 32: (4), 2018

Deep learning algorithms can predict operator actions based on alarm sequences. We developed sequence to sequence learning-based event prediction tools that can be used to for alarm suppression and support the work of operators.

- G. Dörgő, P. Pigler , M. Haragovics, J Abonyi: Learning operation strategies from alarm management systems by temporal pattern mining and deep learning, 28th European Symposium on Computer-Aided Process Engineering (ESCAPE), in Computer Aided Chemical Engineering book series, 2018
- G Dorgo, J Abonyi, Learning and predicting operation strategies by sequence mining and deep learning, Computers & Chemical Engineering 128, 174-187

Process and sequence mining algorithms are useful tools in managing process alarms. As we planned, we studied the applicability of sequence and process mining algorithms in chemical process engineering. We developed an efficient algorithm for alarm management. Even in a case of a simple failure, modern process control systems can cause a vast number of alarms. Due to the overload of the operators these alarm floods may result in tragedical accidents. Alarm management systems can suppress correlated and predictable alarms to reduce the workload of the operators. Since the process units of complex production systems are strongly interconnected, the signals defined on different process variables generate complex multi-temporal patterns. We proposed a multi-temporal sequence mining based approach to extract these patterns and form alarm suppression rules. We demonstrated the applicability of the concept in a vinyl-acetate production technology. The results illustrate the multi-temporal analysis of events defined on process variables can detect causes of alarm, and prevent alarm floods by pro-actively suppressing alarms based on the extracted sequences of events.

- Richard B. Karoly, Janos Abonyi: Multi-temporal sequential pattern mining based improvement of alarm management systems, IEEE Systems Man and Cybernetics Conference, Budapest, 2153, 1-8, 2016
- Gyula Dorgo,Janos Abonyi: Sequence Mining based Alarm Suppression, IEEE ACCESS 6: pp. 15365-15379., 2018

Sequence mining algorithms can be further developed to handle the hierarchical structure of the technologies.

- Dörgő Gy,Varga K,Abonyi J: Hierarchical frequent sequence mining algorithm for the analysis of alarm cascades in chemical processes, IEEE ACCESS 2018: pp. 1-20., 2018

The proposed methods can reduce operator workload.

- Dörgő Gyula,Varga Kristóf,Haragovics Máté, Szabó Tibor,Abonyi János: Towards Operator 4.0, Increasing Production Efficiency and Reducing Operator Workload by Process Mining of Alarm Data, CHEMICAL ENGINEERING TRANSACTIONS 70: pp. 829-834., 2018

Structural analysis and optimisation of complex systems

https://www.abonyilab.com/network-science/structural-analysis

Since we are interested in the data-driven analysis of complex systems, -our attention is shifted to structural analysis. We worked out process mining- and network analysis–based methods for production flow analysis. Production flow analysis includes various families of components and groups of machines. Machine-part cell formation means the optimal design of manufacturing cells consisting of similar machines producing similar products from a similar set of components. Most of the algorithms reorders of the machine-part incidence matrix. We generalized this classical concept to handle more than two elements of the production process (e.g. machine - part - product - resource - operator). The application of this extended concept requires an efficient optimization algorithm for the simultaneous grouping these elements. For this purpose, we proposed a novel co-clustering technique based on crossing minimization of layered bipartite graphs. The presented method has been implemented as a MATLAB toolbox. The efficiency of the proposed approach and developed tools was demonstrated by realistic case studies. The log-linear scalability of the algorithm is proven theoretically and experimentally.

- Csaba Pigler, Ágnes Fogarassy-Vathy, János Abonyi: Scalable co-Clustering using a Crossing Minimization – Application to Production Flow Analysis, Acta Polytechnica Hungarica Vol. 13, No. 2, 209-222, 2016, 2016

We also studied how complex production systems can be optimised by network analysis:

- Tamás Ruppert,Gergely Honti,János Abonyi: Multilayer Network-Based Production Flow Analysis, COMPLEXITY 2018:, 2018

As process mining and event analysis were in the focus of our research, we studied how the developed model can be used to identify operator activities:

- Tamas Ruppert,Janos Abonyi: Software Sensor for Activity-Time Monitoring and Fault Detection in Production Lines, SENSORS 18: (7), 2018

We developed a bipartite network model of education to work transition and a graph configuration model based metric. We studied the career paths of 15 thousand Hungarian students based on the integrated database of the National Tax Administration, the National Health Insurance Fund, and the higher education information system of the Hungarian Government. A brief analysis of gender pay gap and the spatial distribution of over-education is presented to demonstrate the background of the research and the resulted open dataset. We highlighted the hierarchical and clustered structure of the career paths based on the multi-resolution analysis of the graph modularity. The results of the cluster analysis can support policymakers to fine-tune the fragmented program structure of higher education.

- L Gadar, J Abonyi, Graph configuration model based evaluation of the education-occupation match, PloS one 13 (3)

We implemented the above-mentioned algorithm in R

- https://github.com/abonyilab/Edu_Mine_Graph

We worked out a method for modularity analysis and applied to evaluate attractivenes of economic regions:

- Gadar Laszlo,Kosztyan Zsolt T.,Abonyi Janos: The Settlement Structure Is Reflected in Personal Investments: Distance-Dependent Network Modularity-Based Measurement of Regional Attractiveness, COMPLEXITY 2018:, 2018

Since sustainability is critical issues nowadays, we adopted our methodology to evaluate models of complex ecological systems and interlinkages of sustanability issues. https://www.abonyilab.com/systems-engineering/sustainability-science

- Dörgo G.,Sebestyén V.,Abonyi J.: Evaluating the interconnectedness of the sustainable development goals based on the causality analysis of sustainability indicators, SUSTAINABILITY 10: (10), 2018
- Sebestyén V., Bulla M., Rédey Á., Abonyi J.: "Network Model-Based Analysis of the Goals, Targets and Indicators of Sustainable Development for Strategic Environmental Assessment", Journal of Environmental Management, 2019, 238, 126-135

As the complexity of sustainability-related problems increases, it is more and more difficult to understand the related models. Although tremendous models are published recently, their automated structural analysis is still absent. We developed a methodology to structure and visualise the information content of these models. The novelty of the presented approach is the development of a network analysis-based tool for modellers to measure the importance of variables, identify structural modules in the models and measure the complexity of the created model, and thus enabling the comparison of different models. The results highlighted that with the help of the developed method the experts can highlight the most critical variables of sustainability problems (like arable land in the Word 3 model) and can determine how these variables are clustered and interconnected (e.g. the population and fertility are key drivers of global processes). The developed software tools and the resulted networks are all available online.

- Honti G., Dörgő Gy., Abonyi J.: „Review and structural analysis of system dynamics models in sustainability science”, Journal of Cleaner Production, Volume 240, 2019, 118015

Network theory based controllability and observability analysis
https://www.abonyilab.com/network-science/controlability-and-observability

Network theory based controllability and observability analysis have become a widely used technique. We realised that most applications are not related to dynamical systems, and mainly the physical topologies of the systems are analysed without deeper considerations. We drawed attention to the usage of edges defined by the functional relationships among the state variables. The resulting networks differ from physical topologies of the systems and describe more accurately the dynamics of the conservation of mass, momentum and energy. We defined the typical connection types and highlight how the reinterpreted topologies change the number of the necessary sensors and actuators in benchmark networks widely studied in the literature. Based on this concept we worked out several novel algorithms for the analysis of dynamical systems and optmal sensor placement and we worked out a MATLAB toolbox to support the applicability of the methods. According to the google scholar these works are well-cited:

Controllability and observability in complex networks–the effect of connection types D Leitold, Á Vathy-Fogarassy, J Abonyi, Scientific reports 7 (1), 1-9	26	2017
Network distance-based simulated annealing and fuzzy clustering for sensor placement ensuring observability and minimal relative degree D Leitold, A Vathy-Fogarassy, J Abonyi, Sensors 18 (9), 3096	10	2018
Evaluation of the complexity, controllability and observability of heat exchanger networks based on structural analysis of network representations D Leitold, A Vathy-Fogarassy, J Abonyi, Energies 12 (3), 513	9	2019
Design-oriented structural controllability and observability analysis of heat exchanger networks D Leitold, A Vathy-Fogarassy, J Abonyi, Chemical Engineering Transactions 70, 595-600	3	2018
Network-based Observability and Controllability Analysis of Dynamical Systems: the NOCAD toolbox D Leitold, Á Vathy-Fogarassy, J Abonyi, F1000Research 8	1	2019

The results of this pilar of the project are summarized in our book:

https://link.springer.com/book/10.1007/978-3-030-36472-4

Based on this concept we developed a MATLAB toolbox for the structural analysis of complex systems.

- https://github.com/abonyilab/NOCAD

Analysis and optimisation of bussiness processes – stochastic activity time models

https://www.abonyilab.com/optimization/stochastic-systems

At the beginning of the project process mining was a relatively new tool developed for revealing hidden information of log files. Thanks to the quick evolution of this research area and its diversified application, relevant improvements can be observed in several workflow and business models, in which log files have a significant role. We investigated the sequence of sub-processes where the duration and the outcome of the process activities are stochastic so that the process can be successful or failed. We presented a novel methodology for applying survival analysis to find the most relevant components which influence the overall behavior of the process. The survival analysis approach was used to determine the right sub-process order. The proposed decision support methodology can be used to support cost reduction projects. An illustrative example related to a testing process showed that with the help of the developed tools we can determine which activities can be shorter or can be eliminated from the overall test sequence.

- Baumgartner, J., Süle, Z., Bertók, B., & Abonyi, J. (2018). Test-sequence optimisation by survival analysis. Central European Journal of Operations Research, 1-19.

We extended this method for P-graph-based multi-objective risk analysis and redundancy allocation in safety-critical energy systems. As most of the energy production and transformation processes are safety-critical, it is vital to develop tools that support the analysis and minimisation of their reliability-related risks. The resultant optimisation problem should reflect the structure of the process which requires the utilisation of flexible and problem-relevant models. This paper highlights that P-graphs extended by logical condition units can be transformed into reliability block diagrams, and based on the cut and path sets of the graph a polynomial risk model can be extracted which opens up new opportunities for the definition optimisation problems related to reliability redundancy allocation. A novel multi-objective optimisation based method has been developed to evaluate the criticality of the units and subsystems. The applicability of the proposed method eas demonstrated using a real-life case study related to a reforming reaction system. The results highlighted that P-graphs can serve as an interface between process flow diagrams and polynomial risk models and the developed tool can improve the reliability of energy systems in retrofitting projects.

- Süle Z., Baumgartner J., Dörgő Gy., Abonyi J.: "P-graph-based multi-objective risk analysis and redundancy allocation in safety-critical energy systems", Energy (2019), vol. 179, 989-1003.
- Sule, Zoltan, Janos Baumgartner, and Janos Abonyi. "Reliability-Redundancy Allocation in Process Graphs." Chemical Engineering Transactions 70 (2018): 991-996.

The method has been exteneed to empirical working time distribution-based line balancing with integrated simulated annealing and dynamic programming. According to the Industry 4.0 paradigms, the balancing of stochastic production lines requires easily implementable, flexible and robust tools for task to workstations assignment. An algorithm that calculates the performance indicators of the production line based on the convolution of the empirical density distribution functions of the working times and applies dynamic programming to assign tasks to the workstations is proposed. The sequence of tasks is optimised by an outer simulated annealing loop that operates on the set of interchangeable task-pairs extracted from the precedence graph of the task-ordering constraints. Eight line-balancing problems were studied and the results by Monte Carlo simulations were validated to demonstrate the applicability of the algorithm. The results confirmed that our methodology does not just provide optimal solutions, but it is an excellent tool in terms of the sensitivity analysis of stochastic production lines.

- Daniel Leitold,Agnes Vathy-Fogarassy, Janos Abonyi: Empirical working time distribution-based line balancing with integrated simulated annealing and dynamic programming, CENTRAL EUROPEAN JOURNAL OF OPERATIONS RESEARCH 26: pp. 1-19., 2018

We also studied how the developed metholdology can be applied in manufacturing:

- D Leitold, A Vathy-Fogarassy, K Varga, J Abonyi, RFID-based task time analysis for shop floor optimization, 2018 IEEE International Conference on Future IoT Technologies (Future IoT), 1-6
- Tamas Ruppert, Szilard Jaksó, Tibor Holczinger, János Abonyi, Enabling Technologies for Operator 4.0: A Survey, September 2018Applied Sciences 8(1650), Cited: 25 times

Applications in chemistry and bioinformatics research

We also studied how problems in the field of chemistry and bioinformatics can be solved by data-driven approaches. In the first year, we worked out a heuristic algorithm for multiobjective nonlinear optimisation.

The search for compounds exhibiting desired physical and chemical properties is an essential, yet complex problem in the chemical, petrochemical, and pharmaceutical industries. During the formulation of this optimization-based design problem two tasks must be taken into consideration: the automated generation of feasible molecular structures and the estimation of macroscopic properties based on the resultant structures. For this structural characteristic-based property prediction task numerous methods are available. However, the inverse problem, the design of a chemical compound exhibiting a set of desired properties from a given set of fragments is not so well studied. Since in general design problems molecular structures exhibiting several and sometimes conflicting properties should be optimised, we proposed a methodology based on the modification of the multi-objective Non-dominated Sorting Genetic Algorithm-II (NSGA-II). The originally huge chemical search space is conveniently described by the Joback estimation method. The efficiency of the algorithm was enhanced by soft and hard structural constraints, which expedite the search for feasible molecules. These constraints are related to the number of available groups (fragments), the octet rule and the validity of the branches in the molecule. These constraints were also used to introduce a special genetic operator that improves the individuals of the populations to ensure the estimation of the properties is based on only reliable structures. The applicability of the proposed method was tested on several benchmark problems.

The results were published in:

- GYULA DÖRGŐ, JÁNOS ABONYI: GROUP CONTRIBUTION METHOD-BASED MULTI-OBJECTIVE EVOLUTIONARY MOLECULAR DESIGN, HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY, Vol. 44(1) pp. 39–49 (2016)
- G Dörgő, J Abonyi, Hierarchical Representation Based Constrained Multi-objective Evolutionary Optimisation of Molecular Structures, Periodica Polytechnica Chemical Engineering 63 (1), 210-225

All of our results are reproducible as we published our papers and MATLAB/Python programs

- at the website of our research group: http://www.abonyilab.com
- at Researchgate: https://www.researchgate.net/profile/Janos_Abonyi
- and at https://github.com/abonyilab