Towards Predictive Behavior Analysis for Smart Environments

Agnes Koschmider[1], Stefanie Speidel[2]

Abstract: Predictive behavior analysis allows prediction of the (human) behavior based on the analysis of historical data. Efficient approaches for predictive behavior analysis are available for scenarios with structured processes (e.g., based on ERP systems). The prediction of behavior becomes an obstacle when unstructured (decision making) processes underlie the scenario. Scenarios with unstructured processes can be found in smart environments logging sensor (event) streams such as e.g., Smart Home or Connected Cars. No efficient solutions exist to identify abnormal behavior (anomalies) in such smart environments. To provide a solution for anomaly detection in unstructured processes we suggest crossing process engineering with deep learning. Methods from process engineering allow identifying deviations while deep learning improves the robustness of anomalie detection and prediction. This conjunction is a promising approach in order to find an efficient solution.

Keywords: data, process mining, behavior analysis, deep learning

1 Introduction

Predictive behavior analysis allows prediction of the (human) behavior based on the analysis of historical data [Bi55]. Particularly, predictive behavior analysis intends finding patterns, which allow to identify deviations of human behavior and to give predictions how likely it is that activities occur in the future. Efficient approaches for predictive behavior are available for scenarios with structured processes. Blue Yonder[3] is the leading predictive application for structured processes (i.e., they evaluate operating SCM, ERP or HR systems). The forecasts of Blue Yonder are highly robust because it is assumed that human behavior only rarely changes. The prediction of behavior becomes an obstacle when unstructured (decision making) processes underlie the scenario. Unstructured processes are characterized by activities that take place spontaneously and thus the prediction of an appropriate order of activities is hampered. Scenarios with unstructured processes can be found in smart environments logging sensor (event) streams such as e.g., Smart Home or Connected Cars. No efficient solutions exist to

[1] Karlsruhe Institute of Technology, Institute AIFB, 76128 Karlsruhe, agnes.koschmider@kit.edu
[2] Karlsruhe Institute of Technology, Institute AIFB, 76128 Karlsruhe, stefanie.speidel@kit.edu

[3] http://www.blue-yonder.com/

identify abnormal behavior (anomalies) in such smart environments. For instance, a human-being does not necessarily show an abnormal behavior when leaving home directly after getting up while leaving home the day before after breakfast. Predictive behavior analysis might be beneficial for the following scenarios:

- Smart Home: Activities of daily life can be supported by techniques of Ambient Assisted Living. The Smart Home is equipped with numerous sensors (e.g., motion, temperature, heating) in order to provide assistance. Predictive behavior analysis for Ambient Assisted Living might determine the probability of falls or the change of care levels.

- Connected Cars are surrounded by sensors, which track every movement of the driver. Predictive behavior analysis allows an individual communication with the driver warning him/her about personal perilous situations (e.g., feeling uncomfortable when crossing a bridge).

- Industry 4.0 / Industrial Internet: predictive behavior analysis might improve techniques for predictive maintenance (determing the default risk of machines), which are mainly based on heuristics.

The variety of systems helping decision makings of human-beings in smart environments are devoted to a situation-aware assistance. For instance, a Smart Watch that saves biometric profiles and gives recommendations to rest in case that an abnormal blood pressure arises. These approaches are mainly rule-based.

A prerequisite to provide a solution for anomaly detection in unstructured processes is the understanding how unstructured processes are generated from smart environments as shown in Fig. 1. The unstructured (personal) process is derived from traces using process mining algorithms.

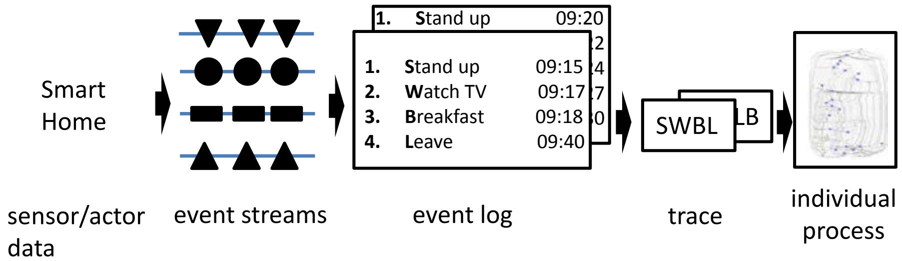

Fig. 1.: From sensor streams to unstructured processes

Identifying anomalies in unstructured processes (called spaghetti processes) is not feasible. It has been shown that multiple graph comparison (a process is a directed, bipartite graph) is an NP-complete problem [Co77]. Therefore, usually the heuristic miner is used to understand the process, which abstracts from exception behavior [WeRi11]. Due to the reduced consideration of exceptional activities, the heuristic miner is therefore not suitable for our purpose. Also a multiple comparison of traces is a NP-complete problem [WaJi94]. Therefore algorithms for pairwise sequence comparison have been proposed in the bio-informatics. The approach of [BoAa12] applies such algorithm in order to detect deviations between traces. With this approach missing or additional activities can be found (e.g., with regard to Smart Home it can be found that a person did not leave home at present day). The identification of missing or additional activities is not enough to identify anomalies. More sophisticated approaches are necessary.

2 Approach

Our approach is based on the crossing of interfield knowledge. Particularly, domain knowledge from process engineering is crossed with deep learning. First, sensor data must be accessible and must be processed in order to derive activities, e.g., moving at home, leaving home. We have access to a large data set of sensors of 50 households that was tracked for a period of one year. Additionally, minutes were taken about the state of the persons (how did they feel). We intend to implement an activity generator allowing to abstract activities from the data / event streams and to create the gold standard. Subsequently, approaches from process engineering and deep learning are applied in order to find anomalies.

- **Process engineering** methods contribute to the preparation of the trace is a way that also structural changes in the behavior of activities can be detected. Changes in the execution of activities allow to identify e.g., activities that are performed in sequence while they have been executed in parallel in the past indicating a performance reduction of the human-being. Abnormal process fragments could be simulated in order to identify temporal changes (e.g., the time to perform a task increases continuously). The view on the problem from a process engineering angle also allows to apply graph matching algorithms in order to detect anomalies in *short* process fragments (e.g., the behavior before and after jogging).

- **Deep learning / machine learning** approaches enable the detection of anomalies and the prediction of behaviour based on historical data. A specific challenge here are time series with heterogenous data that have to be analyzed. Deep Learning, especially recurrent neural networks like long short-term memory (LSTM) is a promising technique for predictive analysis of such data. The advantage of such networks is the ability to process time series with an arbitrary length and the ability to remember relationships in different points in time. It can either be applied on low-level sensor information or on activities derived from sensor data.

3 Challenges

An appropriate crossing of methods from both domains remains a challenge of our approach. Furthermore, a solution for predictive behavior analysis in unstructured processes has to deal with challenges such as (1) Obtaining the gold standard, which shows the intended behavior, (2) Obtaining activities from (sensor) data. Initial solutions have been suggested [SRGM16], (3) Integration of sensor data (e.g., temperature) and activities (moving in the kitchen) during simulation, (4) Obtaining enough data for machine learning approaches, (5) Consideration of (external) environmental influences (e.g., delays of public transportation), (6) Consideration of personal (daily) purposes.

Acknowledgement: The idea of the project has been elaborated together with Melanie Herschel from the University of Stuttgart.

4 References

[Bi55] Bieri, J.: Cognitive complexity-simplicity and predictive behavior. The Journal of Abnormal and Social Psychology, Vol 51(2), 1955, pp. 263-268.

[BoAa12] Jagadeesh Chandra Bose, R. P.; van der Aalst, Wil M. P.: Process diagnostics using trace alignment: Opportunities, issues, and challenges. Inf. Syst. 37(2): 117-141, 2012

[Co77] Cook, S. A.: The complexity of theorem-proving procedures, Proc. 3rd ACM Symposium on Theory of Computing, 1977, pp. 151–158

[SRGM16] Senderovich, A.; Rogge-Solti, A.; Gal, A.; Mendling, J.; Mandelbaum, A.: The ROAD from Sensor Data to Process Instances via Interaction Mining. CAiSE 2016: 257-273

[WaJi94] Wang, L., Jiang, T.: On the complexity of multiple sequence alignment. Journal of Computational Biology 1(4), 1994, pp. 337–348

[WeRi11] Weijters, A.J.M.M.; Ribeiro, J.T.S.: Flexible Heuristics Miner (FHM). CIDM 2011: 310-317