Extraction of Qualitative Models for Cyber-Physical Systems

Short paper

Balázs Márk Hain

hainb@edu.bme.hu
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Budapest, Hungary

András Pataricza

pataricza.andras@vik.bme.hu
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Budapest, Hungary

***Abstract*— The paper focuses on interpretable and explainable discrete qualitative modeling of observations. The core abstraction problem in qualitative model creation is the clustering of continuous observed data belonging to the same operational domain exposing a similar behavior into a single qualitative state. The structure of qualitative models well-reflect the dynamic architecture of the target system. Thus, they are natural choices to build a digital twin serving as the core for supervisory control of complex cyber-physical systems.**

The quality of clustering crucially influences the faithfulness of the qualitative model and consequently the efficiency of the control. Moreover, proper detection and management of outliers is a primary objective in critical applications.

Our research proposes a technique that combines dimensionality reduction and clustering methods that accurately separate the operational domains while also recognizing outliers using well-fitting cluster borders. This solution also features various interpretability methods that can showcase the cohesive factors amongst the operating regions and guide the understanding of the functionality as well.

I. Introduction

A. Relation with CPS applications

The concept of "Digital Twin" (DT) has been heavily integrated into operation-critical systems, in industrial automation of CPS applications, to maintain a high level of overall monitoring and control of the individual processes. [10] DT is a comprehensive representation of a complex physical system, which is able to create high-fidelity models, to enclose the dynamics both of the environment and the distributed system.

Complex CPS systems frequently integrate COTS components and external services. The modeling of their (extra)functional properties frequently needs empirical system identification complementing the priory known design information, like the engineering design model. [9] That reflects and differentiates distinctive operating states while also having the ability to obtain the system's expected behavior and detecting anomalies and outliers.

B. Qualitative modeling

Qualitative modeling expresses conceptual knowledge about the system structure, dynamics, causality relations in its functionalities, assumptions about its operation, and qualitatively distinct operational domains. Common engineering thinking motivates this modeling paradigm, resulting in easy-to-understand and well-interpretable models.

Its core idea is to extract an abstract representation from detailed (partial) models and observations by aggregating and discretizing all the continuous features. Qualitative abstraction maps entire subsets of the continuous state space corresponding to an operational domain exposing similar behavior to a single qualitative state. Similarly, ordinal qualitative values represent subdomains of individual continuous variables (e.g., *low, medium, high*) into ordinal variables, thus preserving their relative magnitude and (partial) ordering. [7]

Qualitative modeling has a long tradition in different fields of science as descriptive means; however, its use for CPS control raises specific challenges: (1) The quality of the model must be guaranteed as the decisive factor of its faithfulness; (2) Insufficient coverage of potentially dangerous operational domains may lead to hazardous control errors; (3) Outliers need special care in critical applications.

The derived clusters need a thorough V&V by statistical and engineering expert analysis to cope with these demands. The latter needs the interpretability and explainability of the models. Our primary research problem was assessing the appropriateness of the rapidly growing repertoire of xAI (eXplainable Artificial Intelligence) methods for the V&V of qualitative models. xAI explainers serve to highlight the main features of data sets and models derived from them.

II. Qualitative model extraction

A. Problem description

The pilot example originates in an experiment evaluating remote web-service by measuring the latencies in different phases in a client-server task invocation (communication: Response Time [*RT*], data processing: Request Processing Time [*RPT*], total: Round Trip Time [*RTT*]). Experiments were concluded homogeneously across several spatial and temporal locations. The timestamped [*start.time*] logs capture geographic properties, such as *location*, *country*, time zones [*time*]; client attributes: [*client.type*], IP address (*IP*), data center [*DC*]. The dependence of the total service time (*RTT*) on individual factors outlines the central issue of remote data-processing involved in the given typical CPS application.

An initial partitioning by visual exploratory data analysis identified cases (*normal, border class, faulty*) (Fig. 1a) by

(a) Class distribution of the initial data

(b) Clustering (10 clusters)

(c) Back-annotation of the clusters to the data

Fig. 1: Clustering process

the qualitative discretization of the RT values (*low, medium, high*); and the next step is the estimation of dependencies on input factors by partitioning the clusters.

B. Overview of the approach

Implementations use a wide variety of sensors both for the cyber and the physical parts of the CPS to assure proper observability of the control and controlled systems, thus resulting in a many-dimensional stream of big data. The extraction of a manageable and interpretable qualitative model and supervisory control demands selecting a limited subset of core variables (feature selection).

Our approach follows the following process. Firstly, a transformation of the entire dataset to a compact and understandable 2D visual representation (Fig. 1b). It exposes the potential operation domains, which are impossible to reveal in the original view (Fig. 1c). Statistical and inference methods audit the correspondence of the indicated candidate clustering and the distinct operational domains of the system. Statistics evaluate the quality of separation, and any noise introduced, forming a solid foundation for feature engineering. Reasoning explores the cohesive factors amongst each qualitative operating state. It provides insight into their inner functionality. If complemented with variable importance metrics, it highlights the contribution of different technical factors in the individual domains.

C. Dimension reduction and clustering

Dimension reduction (DR) projects a set of observations into a low-dimensional space while preserving most of the knowledge on the input features, like the cluster structure in the high-dimensional space. It deals with the spatial disposition of the samples, i.e. whether other points of data should be close or far away from the selected one (local & global structure). [1]

Traditional DR methods, like PCA, account for only the linear separation of the data, which is an excellent first approach for rough modeling but inadequate for generating complex models. Advanced techniques (like t-SNE, UMap) tend only to affirm the primary model's local or global structure. A new algorithm, PaCMAP [11] combining mid-near pairs of observations and dynamic graph optimization, performed well in our benchmark on both local and global structure (cf. Fig. 1b).

Modeling physical objects raises additional requirements of representational invariance for modeling physical systems: (i) results must be independent of the unit of measurement and (ii) substituting a set of features with another one having the same information content (like transformed metrics) should not affect the results. Since none of the algorithms mentioned above fulfills these in their basic form, we propose additional solutions. As only the relative magnitude of a particular feature is essential in qualitative modeling, feature scaling helps: e.g., domains of each variable should be normalized according to the mean of their distribution. The second problem can be solved by careful feature selection, discussed in a later chapter. Another known insufficiency of DR methods is their inability to deal with categorical data. We suggest resolving this issue by 1-of-N coding similarly as machine learning algorithms do it.

Any commonly used algorithm can cluster the transformed data. In this research, DBScan [6] (a fast, density-based clustering) was used, as it performs the best with PaCMAP's output. A further significant advantage of DBScan is the recognition and marking of outliers, instead of only suppressing them.

III. Model analysis, validation, and verification

A. Statistical evaluation and feature selection

The quality of the clustering outcome heavily depends on selecting relevant descriptor variables to avoid the underfitting of the model or redundancy-induced surplus complexity. In this scenario, interactive partitioning-based clustering is preferred since the usual/automatic methods are less helpful. Interactive clustering tailors the process to specific application domains validated by several evaluation measures. [3] In a purely unsupervised learning case, correlation metrics (e.g., phi(K) [2] - which can process categorical features as well) can be used to select the relevant features.

A target variable can be constructed for the given problem using prior knowledge about the system extracted from the engineering model. The proposed iterative feature selection process is similar to a concept called Minimum Redundancy Maximum Relevance - mRMR [12]. This idea seeks to identify a small subset of features that collectively have the maximum possible predictive power while minimizing the size of the collection, omitting the redundant variables. The field of xAI offers a Feature Importance that showcases the contribution of each individual feature to the target metrics. Therefore an ordered list of variables ranked by their respective importance guides the selection of the most relevant

features. Fundamental means of cluster evaluation, such as noise level, the ratio of homogeneity, and completeness, support the V&V of clusters.

Firstly the input attributes are sorted based on relevance metrics, such as the Feature Importance value delivered by an xAI algorithm. Then, a feature is selected with high relevance in each iteration, starting from the most significant feature. The selection must stop when the noise level rises, indicating the irrelevance of the additional variables regarding the initial problem.

B. Analysis of qualitative domains

Operational domain identification aims to identify homogeneous domains composed of data points of similar system behavior. The clustering process already determined these domains phenomenologically; however, discovering their boundaries and interpreting their inner functionality is still a primary intention.

xAI is a field that provides various means to explain complex Machine Learning (ML) models by making them directly interpretable. Our approach aligns with the current ongoing trend to shift from the extensive domain knowledge, (computationally heavy) construction, and fitting of models towards the V&V and interpretation of the model. We believe that introducing increasing (semi-)automation to the modeling part leads to a better understanding of the analyzed data.

As xAI is currently a rising field, many open-source explanation generators are available. The paper proposes two xAI-based distinct methods for this purpose.

- The first one creates to each cluster an approximate **symbolic membership function**. Here AIX360 can be directly applied, as it specializes in constructing a directly interpretable model based on data, either the initial dataset or ML-generated model of it.
- The second one uses DALEX to estimate **explanatory metrics and visualizations** highlighting the main characteristics of the domain after fitting an arbitrary ML model to it.

Two particular clusters will be the primary subject of the analysis in the pilot example. Let **Cluster A** be the well-separated set of data at the top on Fig. 1b marked with dark blue. The significant curve-shaped green-colored cluster at the bottom of the figure is referred to as cluster **Cluster B**.

1) Membership function of the operation domain via Rule Generation: The first approach fits a logic membership function as an abstract representation of the underlying domain in the form of rule sets, using the Boolean Decision Rules via Column Generation (BRCG) [5] algorithm. Rules are simple, thus easy to understand, and interpretation-friendly, as proven by many popular ML modeling paradigms, like decision trees, random forests. The core idea of BRCG is the composition of the domain model from a set of subdomains, each defined individually by a decision rule. Optimizing the user-weighted sum of the coverage accuracy of the particular domain versus the number and complexity of the decision rules assures a trade-off between the contradicting aspects of the accuracy and the interpretability of the composite Boolean domain model.

Having no prior knowledge of the characteristics of **Cluster A & B**, we analyzed these regions with the BRCG.

Ruleset A		Ruleset B	
RT > 2598.20 RPT > 828.00 ip = 208.87.25.162	region 1	RTT > 1933.00 RPT > 828.00 client = Microsoft	region 1
		RTT > 1933.00 RPT > 656.00 ip = 209.188.85.60 start.time <= 1359777	region 3
RT > 2598.20 RPT > 719.00 start.time <= 1359769 location = Durham	region 2	RTT > 1933.00 RPT > 781.00 ip = 64.20.37.202 start.time > 1359769	region 2
		ip = 67.227.216.114	region 4

Fig. 2: Result of the rule generation for the respective clusters

Results (Fig. 2) indicate that samples with only high latency values are present in the observed domains. These samples originate from specific locations and timeframes. From the rulesets, it is deductible that both clusters represent substandard data - due to the RT value being high - indicating the presence of classes: *border class*, and *faulty*. Cluster A shows higher RT base values than B, however, overlaps may be possible, therefore no further distinction could be made. Additional details (like spatial and temporal properties) will be the subject of root cause analysis.

The membership function approximation approach may provide accurate results, but it only returns a single solution without transparency of the process. For instance, it hides the reason for inclusion or omission of the individual factors into the rules. Moreover, a moderate change in the weights in the cost function may lead to drastically different rule sets.

2) Operating domain characterization: Operating domain characterization aims at the well-interpretable exposition of the cohesion and separation factors and the boundaries of a cluster. It relies on the concept of **prototype generation**. The prototype is a single point representing the characteristics of all points in the entire cluster, if it is sufficiently homogeneous (similarly to an equivalence class leader in automata theory). The ProtoDash [8] implementation suffers from performance issues upon processing a high-dimensional large-scale dataset. However, its core idea can be adapted by using the central element in a density-based cluster as its representative prototype. Note that clusters mapped to qualitative states have to be homogeneous by their definition.

Using a prototype reduces the characterization of a cluster to the local analysis of a single data instance. modelStudio [4] provides local explainability techniques supporting:

- **Extraction of relevant features**: Break Down / Shapley values. These methods determine the features' contribution to the symbolic membership function, using its partial derivative (the sensitivity) to the individual input features. Variables of a marginal influence on the model are subject to the following sensitivity analysis.
- **Sensitivity analysis**: Ceteris Paribus (CP) profiles. CP demonstrates how sensitive the model is for a slight

change in a particular variable, therefore showing partial model responses around a candidate point. CP profiles serve as a mean of **cross-cutting projection** of the observed cluster. CP supports the identification of the cluster cohesion variables and their respective domains.

The above-mentioned analytical steps on Clusters A and B, result in a better insight to their composition.

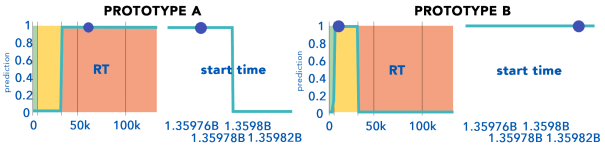

Fig. 3: CP profiles representing Cluster A, B
RT's background signals initial class separation (Fig. 1a)

The highlighted points on the CP profiles are the prototypes of the particular clusters (Fig. 3). Both *RT* and *start.time* contribute to the symbolic membership function in Cluster A, which is sensitive only to the higher values of *RT*'s and the low values of *start.time*. Cluster B is constrained to a specific subset of *RT*, to which the feature demonstrates high sensitivity while being insensitive to the *start.time*.

Comparing the identified domains of sensitivity with the qualitative regions, Cluster A contains only samples that qualify as *faulty data*, this way it fulfils the criterion of homogeneity. At the same time, the samples in this set originate **only** from specific time period *start.time*. Accordingly, Cluster A lies in the cut set of the qualitative domains *RT=faulty* and *start.time=low*. In contrary, Cluster B seems to contain the **entire** class of *border class* samples, as it depends only on RT and the clusters are disjoint. However, V&V indicates overlap with different qualitative domains of *RT* violating the requirement of homogeneity, thus the initial, somewhat arbitrary visual clustering of *RT* was slightly shifted. Hence, the intrinsic capability of checking the assumptions helps to detect even minor mistakes (where after correction the result becomes consistent).

3) Evaluating the results: The clustering method approximately separated the samples according to the initial engineering viewpoint - by the qualitative magnitude of transfer time *RT*.

It is also essential to understand that the clusters formed by the DR method constructed their own qualitative domain itself, thereby eliminating the need for various visual or statistical analyses to construct them. Reprojecting (Fig. 1c) the clusters association with their contained samples refines the initial visual exploratory analysis-based partitioning, and it readjusts the operational region boundaries slightly, assuring a coherent structure of the model.

IV. CONCLUSIONS

The concluded examination reinforces and justifies the use of a qualitative model. The proposed technique showcased its ability to define a qualitative membership function for each operational region.

With the use of xAI methods, cohesive factors and their respective importance can be thoroughly understood. With that knowledge, homogeneity level and the separation quality regarding the constructed clusters are easily observed. That allows for interpretable, verifiable, and validable qualitative model construction.

The introduction of a solid xAI repertoire to the analysis process has shifted the focus towards the V&V phase from the domain knowledge by forming the qualitative regions as a byproduct of the analysis process.

The analysis results related to the boundaries of the different operation domains are directly relevant in setting up the monitoring part of a hybrid supervisory control scheme. Here the detection of entering a new operation domain is detected by comparators watching the crossing of the boundaries and trigger the control actions, referred to as scenario identification in the technical literature.

Further research should examine the question of exhaustive outlier detection, as well as the causal relationships between explained factors and the target variable. Formally proving the correctness of our hypothesis formulated by interpretability metrics is also a key subject of interest.

ACKNOWLEDGMENT

The research was cofounded by the PROTECTME - Protecting Operational Technologies of Medium Enterpises (EIT Digital) project. (No. 21293)

REFERENCES

[1] Shaeela Ayesha, Muhammad Kashif Hanif, and Ramzan Talib. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59:44–58, 2020.

[2] M. Baak, R. Koopman, H. Snoek, and S. Klous. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics Data Analysis*, 152:107043, 2020.

[3] Juhee Bae and et al. Interactive clustering: A comprehensive review. *ACM Comput. Surv.*, 53(1), Feb. 2020.

[4] Przemyslaw Biecek and Tomasz Burzykowski. *Explanatory Model Analysis*. Chapman and Hall/CRC, New York, 2021.

[5] Sanjeeb Dash, Oktay Günlük, and Dennis Wei. Boolean decision rules via column generation. *arXiv*, 1805.09901, 2020.

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96, page 226–231. AAAI Press, 1996.

[7] Kenneth D Forbus. Qualitative modeling. In Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, *Handbook of knowledge representation*, volume 3, pages 361–393. Elsevier, 2008.

[8] Karthik S. Gurumoorthy, Amit Dhurandhar, Guillermo Cecchi, and Charu Aggarwal. Efficient data representation by selecting prototypes with importance weights. *arXiv*, 1707.01212, 2019.

[9] Imre Kocsis, Ágnes Salánki, and András Pataricza. Measurement-based identification of infrastructures for trustworthy cyber-physical systems. In Alexander Romanovsky and Fuyuki Ishikawa, editors, *Trustworthy Cyber-Physical Systems Engineering*, pages 395–420. Chapman and Hall/CRC, 2016.

[10] Fei Tao, Qinglin Qi, Lihui Wang, and A.Y.C. Nee. Digital twins and cyber–physical systems toward smart manufacturing and Industry 4.0, correlation and comparison. *Engineering*, 5(4):653–661, 2019.

[11] Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. *arXiv*, 2012.04456, 2021.

[12] Zhenyu Zhao, Radhika Anand, and Mallory Wang. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In *2019 IEEE Intl. Conf. on Data Science and Advanced Analytics (DSAA)*, pages 442–452, 2019.