

Bericht zur Max-Buchner-Forschungsarbeit

„Self-adjusting Digital Twin for Optimal Operation of (Bio-)Chemical Processes under Uncertainty“ (MBFSt-Kennziffer: 3733)

Dr.-Ing. Erik Esche, Fachgebiet Dynamik und Betrieb technischer Anlagen, Technische Universität Berlin

1. Aufgabenstellung und Zielsetzung

Seit Jahren besteht der Wunsch nach sich kontinuierlich selbst-optimierenden Chemieanlagen, um Energie- und Ressourcen während des Betriebs einzusparen. Der Einsatz von Real-time Optimization (RTO) hierfür ist allerdings nach wie vor sehr gering, da der große Implementierungsaufwand, nötiges Expertenwissen sowie die hohen Kosten immense Hürden darstellen. Seit einigen Jahren werden daten-getriebene Methoden / Techniken diskutiert, um die aufwändig zu erstellenden rigorosen Anlagenmodelle zu ersetzen und die Kosten für den selbst-anpassenden, autonomen Betrieb von (bio-)chemischen Prozessen zu ermöglichen. Eine Herausforderung für den Aufbau verlässlicher daten-getriebener Modelle realer Anlagen liegt in den häufig entscheidenden „seltenen“ Qualitätsmessungen. Während Temperaturen, Drück, Durchflüsse und Füllstände kontinuierlich gemessen werden, finden Qualitätsmessungen häufig nur im Stunden-, Tages- oder Wochentakt statt. Zur Einbeziehung dieser seltenen Messdaten in den Aufbau daten-getriebener Modelle stehen halbüberwachte Lernverfahren zur Verfügung (semi-supervised learning - SSL). Allerdings wurden diese bislang noch nicht auf Zeitserien bzw. für Optimierungsanwendungen in der Verfahrenstechnik eingesetzt. Dementsprechend werden im Rahmen dieses Projektes SSL-Methoden um drei Bausteine erweitert: (1) die Quantifizierung von Unsicherheiten enthalten in daten-getriebenen Modellen, (2) die Adaption von SSL-Methoden für Zeitserien bzw. dynamische Prozesse. (3) die Anwendung auf stark geclusterte Anlagendaten. Nach Abschluss dieses Forschungsvorhabens soll ein Modellierungskonzept für SSL stehen, welches (1) eine hohe Qualität von Eingangs-Ausgangs-Beziehungen garantiert, (2) eine hohe Vorhersagegüte in Bezug auf seltene Qualitätsmessungen aufweist und (3) SSL als Methode für Online-Optimierung unter Unsicherheiten zur Verfügung stellt.

2. Durchgeführter Arbeitsplan

Der gesamte Projektplan ist auf 3 Jahre ausgelegt. In diesem Bericht wird ausschließlich über das erste, bereits abgeschlossene Jahr berichtet. Im ersten Projektjahr standen zwei Arbeitspakete an: (AP1) die daten-getriebene Modellierung und Unsicherheitsmodellierung für SSL; (AP2) die Anwendung und Evaluation auf Basis vorhandener industrieller Anlagendaten. AP1 erfolgt rein auf Basis simulativ erzeugter Anlagendaten, bei welchen die Messfrequenz für Qualitätsmessungen sowie Messfehler / Messrauschen direkt beeinflusst werden können. Auf Basis dynamischer Prozessmodelle werden typische Betriebstrajektorien erzeugt: ein längerer Betriebshorizont mit einigen wenigen Betriebspunktwechseln je nach Anlagenlast. Diese simulativ

erzeugten Daten werden zusätzlich mit Messrauschen und Ausreißern versehen. Darüber hinaus wird über bestimmte Modellparameter gesampelt, um die übliche Variation im Anlagenverhalten darzustellen. Hierzu wurde initial auf zwei bekannte Modelle zurückgegriffen, eine dynamische Implementierung des Williams-Otto-Reaktors sowie das dynamische Modell eines Bioethanolprozesses [3]. Auf Basis der so erzeugten Prozessdaten wird sowohl die datengetriebene Modellierung als auch die Unsicherheitsmodellierung durchgeführt. In Abb. 1 ist das Ablaufschema für die Unsicherheitsmodellierung dargestellt, welches mit Hilfe von scikit-learn [1] implementiert wurde. Die vollständige Implementierung ist in [2] beschrieben.

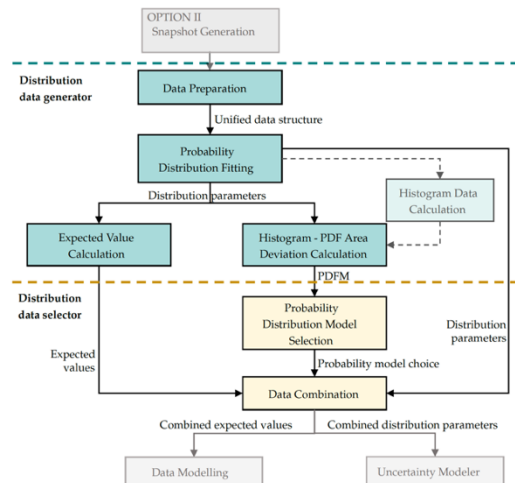


Abbildung 1 - Ablaufschema der Unsicherheitsmodellierung – in [2] veröffentlicht.

Die datengetriebene Modellierung basiert auf einer TensorFlow-Implementierung von Jean et al. [4], welche erweitert wurde, um Zeitserien in den Eingangsgrößen zuzulassen. Hierbei wurden verschiedene Optionen für den Umfang bzw. das Format der Zeitserien in Betracht gezogen. Im Rahmen von AP 2 wurden die Unsicherheitsmodellierung sowie die datengetriebene Modellierung dann auf einen industriellen Chlor-Alkali-Prozess angewendet.

3. Ergebnisse

Abb. 2 zeigt künstliche Betriebsdaten für den dynamischen Williams-Otto-Reaktor. Je nach Prozessgröße, wurden hier einige Sprünge aufgegeben, um verschiedene Betriebspunkte anzufahren. Alle Größen sind skaliert. Auf Basis dieser Anlagendaten wurde dann die Messfrequenz einer Qualitätsmessung im Reaktor von minütlich, stündlich, täglich bis wöchentlich variiert, während alle übrigen Messdaten weiterhin kontinuierlich vorliegen. Der Unterschied zwischen SSL und klassischer Regression liegt darin, dass die Verteilung von Eingangsdatenpunkten ohne Ausgangsdatenpunkte (Labels) ebenfalls beim Training des Regressionsmodells berücksichtigt wird. Hierfür wird in der Zielfunktion des SSL-Trainings ein zusätzlicher Summand eingefügt, der die Varianz der Eingangsdatenpunkte (ohne Labels) beschreibt. Dieser zweite Summand kann beliebig gewichtet werden. Basierend auf den Williams-Otto-Daten sehen wir eine relativ große Sensitivität der Prädiktionsgüte abhängig von der Wahl des Gewichtungsfaktors. Hierbei lag für oben gezeigtes Betriebsszenario die größte Prädiktionsgüte bei der höchsten Gewichtung des Varianzterms vor. Abb. 3 zeigt hierfür eine Prädiktion der Zeitserie im Vergleich zur „Modellwahrheit“, wenn dem Training stündliche Qualitätsmessungen vorliegen über

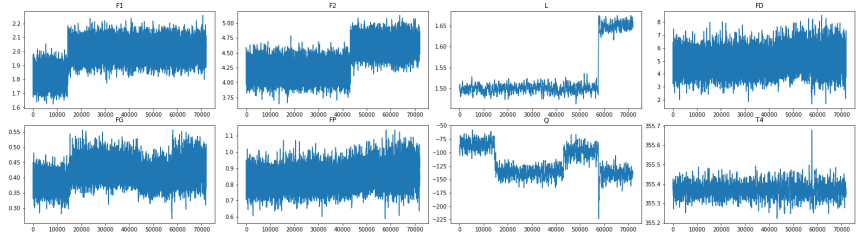


Abbildung 2: Künstlich erzeugte Betriebsdaten für den dynamischen Williams-Otto-Reaktor

Alle Größen sind skaliert. Auf Basis dieser Anlagendaten wurde dann die Messfrequenz einer Qualitätsmessung im Reaktor von minütlich, stündlich, täglich bis wöchentlich variiert, während alle übrigen Messdaten weiterhin kontinuierlich vorliegen. Der Unterschied zwischen SSL und klassischer Regression liegt darin, dass die Verteilung von Eingangsdatenpunkten ohne Ausgangsdatenpunkte (Labels) ebenfalls beim Training des Regressionsmodells berücksichtigt wird. Hierfür wird in der Zielfunktion des SSL-Trainings ein zusätzlicher Summand eingefügt, der die Varianz der Eingangsdatenpunkte (ohne Labels) beschreibt. Dieser zweite Summand kann beliebig gewichtet werden. Basierend auf den Williams-Otto-Daten sehen wir eine relativ große Sensitivität der Prädiktionsgüte abhängig von der Wahl des Gewichtungsfaktors. Hierbei lag für oben gezeigtes Betriebsszenario die größte Prädiktionsgüte bei der höchsten Gewichtung des Varianzterms vor. Abb. 3 zeigt hierfür eine Prädiktion der Zeitserie im Vergleich zur „Modellwahrheit“, wenn dem Training stündliche Qualitätsmessungen vorliegen über

einen Horizont von 70'000 Sekunden. Hierbei wurde auf das Messrauschen verzichtet, um eine Lesbarkeit zu ermöglichen. Trotz der geringen Anzahl an Messpunkten ist die Vorhersagegüte überraschend gut, sowohl Dynamik als auch absolute Konzentrationswerte werden mit geringem Fehler wiedergegeben. Eine breitere Untersuchung für längere Zeithorizonte und weitere Anlagenmodelle steht noch aus. Bei geringerer Messfrequenz (täglich, wöchentlich) sinkt erwartungsgemäß die Vorhersagegüte, die Wichtigkeit des Varianzterms bleibt bestehen.

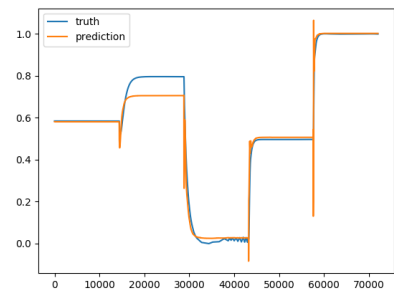


Abbildung 3: Vergleich von Qualitätsprädiktion und Modellwahrheit (ohne Messdatenrauschen, normalisiert)

Parallel zur SSL-Adaption für Zeitserien fand die Quantifizierung und Modellierung der Unsicherheiten statt. Hierbei lag der Schwerpunkt zunächst auf der Identifikation von passenden Verteilungsmodellen. Einen Eindruck der betrachteten Verteilungen gibt Abb. 4, worin die Verteilung „weibull max“ die größte Güte zeigt.

Nach den künstlichen Daten aus AP1 wurden SSL- und Unsicherheitsmodellierung auf industrielle Prozessdaten aus einer Chlor-Alkali-Elektrolyse (CAE) angewendet. Die Ergebnisse hiervon sind teilweise in [2] veröffentlicht. Die Daten der CAE können mit einer hohen Prädiktionsgüte wiedergegeben werden. Für die Beschreibung der Unsicherheiten kommen insbesondere Beta-Verteilungen zum Einsatz, deren Parameter abhängig vom Betriebszustand modelliert werden kann (siehe Abb. 7 in [2]).



Abbildung 4: Vergleich der Unsicherheitsmodellgüte. Je höher der Wert, desto besser ist das Zeilen- im Vergleich zum Spaltenmodell, siehe auch [2].

4. Fazit

Während des ersten Projektjahrs erwies sich SSL bereits als sehr nützliches und auch verlässliches Werkzeug, um seltene Qualitätsdaten in Regressionsmodelle für Chemieanlagen einzubeziehen. Ebenso gelingt es auf Basis von realen Anlagendaten bereits gut die bestehenden Unsicherheiten in den Regressionsmodellen darzustellen. In den folgenden zwei Projektjahren werden diese Methoden nun in den praktischen Einsatz für die kontinuierliche Reoptimierung von Anlagen unter dynamischen Laständerungen und Störungen gebracht. Insbesondere für Anlagen mit selteneren Qualitätsmessungen (täglich bis wöchentlich) wird hier noch eine weitere Unterstützung von SSL durch simulationsgetriebene Modelle erforderlich. Hierzu laufen bereits erste Arbeiten, um die erhaltenen SSL-Modelle zur Korrektur von simulationsbasiert erhaltenen Ersatzmodellen einzusetzen. Ausdrücklich möchte ich mich für die finanzielle Unterstützung der Max-Buchner-Forschungstiftung bedanken, dank derer studentische Hilfskräfte zur Umsetzung der Arbeiten eingestellt werden konnten.

5. Literatur

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., J. Mach. Learn. Res., 12, 2825-2830, 2011
- [2] B. Häussling Löwgren, J. Weigert, E. Esche, J.-U. Repke, Sustainability, 12, 2450, 2020
- [3] S. Ochoa, G. Wozny, J.-U. Repke, 2010, J. Proc. Contr. 20, 983-998
- [4] N. Jean, S.M. Xie, S. Ermon, 2018, arXiv:1805.10407