

# METHODEN DER ERKLÄRBARKEIT VON REINFORCEMENT LEARNING MODELLEN ZUR TOPOLOGIEOPTIMIERUNG IN ÜBERTRAGUNGSNETZEN

Harald JENDRIAN<sup>1</sup>, Reinaldo TONKOSKI<sup>1</sup>, Ilja KRYBUS<sup>2</sup>

## Hintergrund

Der zunehmende Zubau von Windkraft- und Photovoltaikanlagen sowie Großbatteriespeichern erhöht die Komplexität der Netzführung in 380-/220-kV-Übertragungsnetzen. Die daraus resultierenden strom- und spannungsbedingten Engpässe werden üblicherweise durch Redispatch behoben. Alternativ können Schalthandlungen N-1-Befunde beheben oder Redispatchpotential verringern. Die kombinatorische Vielfalt der Schaltgeräte in Umspannwerken erschwert jedoch das Auffinden optimierter Topologien, weshalb diese in der Netzeinsatzplanung oft unzureichend berücksichtigt werden [1].

Dieser Beitrag stellt einen Reinforcement Learning Ansatz zur Topologieoptimierung für ein deutsches Übertragungsnetz vor (max. Erzeugungsleistung: 4,6 GW, Jahreshöchstlast: 7,7 GW). Ein Agent interagiert mit einer *Environment* und erhält als *Observation Space* u.a. Betriebsspannungen, Leitungsauslastungen sowie Einspeise- und Lastzeitreihen. Das Optimierungsziel (*Reward*) ist die Minimierung der maximalen Leitungsauslastung; Konvergenzfehler und spannungslose Knoten werden durch *Penalties* bestraft. Der *Action Space* umfasst das diskrete Betätigen von Leistungs- und Trennschaltern. Um praxisnahe Schaltzustände abzubilden, werden Schalter als *Grouping Actions* zusammengefasst; unzulässige Handlungen – etwa das gleichzeitige Belegen einer Leitung auf zwei Sammelschienen – werden durch *Action Masking* unterbunden. So ermöglicht der Agent die optimale Kombination von Ein-/Ausschaltungen von Leitungen, das Betätigen von Querkupplungen und das Durchführen von Sammelschienenwechseln in den Umspannwerken. Zum Einsatz kommt ein *maskable Proximal Policy Optimization* (maskablePPO) Algorithmus, der iterativ Policy und Value Function aktualisiert. Die Implementierung basiert u.a. auf Simbench Netzmodellen, pandapower, stable-baselines3, gymnasium und TensorBoard. Der Anwendungsfall adressiert eine vorausschauende Netzeinsatzplanung, in der topologische Massnahmen für den Folgetag bestimmt werden – etwa als Non-costly Remedial Actions im Rahmen europäischer Day-Ahead Congestion Forecast (DACF) Prozesse. Abbildung 1 zeigt einen beispielhaften Dezembertag, an dem 12–20 Querkupplungen geöffnet und 12–27 Sammelschienenwechsel vorgeschlagen werden; in den frühen Abendstunden empfiehlt der Agent zusätzlich das kurzfristige Ausschalten einer Leitung. Dadurch sinkt die durchschnittliche maximale Netzauslastung von 108,99 % auf 85,57 %. Zudem zeigt der Tagesverlauf, dass diese Maßnahmen zwischen 00:30 und 5:30 Uhr N-1-Befunde beseitigen und Redispatchabrufe vermeiden.

DAYAHEAD OPERATIONAL SCHEDULE													
Topology Options	Indicator	00:30	01:00	01:30	02:00	02:30	03:00	03:30	04:00	04:30	05:00	05:30	(...)
Reference Topology	N-1 Max line loading (before)	135.0	133.4	159.7	128.1	134.4	136.1	157.3	129.3	131.9	132.4	137.1	
Agent 1	N-1 Max line loading (after)	94.0	140.4	118.2	93.2	152.9	86.7	129.3	96.9	100.6	152.4	110.5	
	Reduction rate	30.4	-5.3	26.0	27.3	-13.7	36.3	17.8	25.1	23.7	-15.1	19.4	
	#NodeSplitting	16	17	19	18	14	17	15	18	17	15	16	
	#BusAssignment	18	20	19	21	20	23	20	24	17	19	13	
	#LineDisconnect	0	0	0	0	0	0	0	0	0	0	0	
	#LineConnect	0	0	0	0	0	0	0	0	0	0	0	

Abbildung 1: Tabellarische Übersicht über den Einsatzplan der topologischen Massnahmen 25.12.

<sup>1</sup> Technische Universität München, harald.jendrian@tum.de, reinaldo.tonkoski@tum.de

<sup>2</sup> BearingPoint GmbH, ilja.krybus@bearingpoint.com

Jede Topologieoption wird mit einer Ausfallvariantenrechnung auf ihre Wirksamkeit hin überprüft – es fällt auf, dass Netzsituationen entstehen können, die zwar zu einer N-0 Reduktion führen, jedoch in vermaschten Netzen nicht unbedingt zu einer N-1 Reduktion beitragen. Um Entscheidungen besser nachvollziehen zu können und um Vertrauen in den Agenten zu schaffen, wird in diesem Beitrag die Netzsicherheitsrechnung, um Methoden der Erklärbarkeit (*Explainable AI, XAI*) erweitert.

## Methodik

Mithilfe von *Shapley Additive exPlanations* (SHAP) lässt sich das Verhalten des *maskablePPO* Agenten analysieren, um den Entscheidungsprozess erklären zu können. Für jedes Eingabemerkmals (*Feature*) wird dessen marginaler Beitrag zur Modellvorhersage (*Probability*) quantifiziert, indem systematisch berechnet wird, wie sich die Ausgabe bei An- oder Abwesenheit dieses Merkmals verändert [2]. Um die Berechnungskomplexität im Beobachtungsraum (*Observation*) zu reduzieren, kommt die *KernelSHAP* Approximation in Kombination mit einer Gruppierung semantisch zusammengehöriger Merkmale zum Einsatz. Anhand der SHAP Werte lässt sich beschreiben, welche Merkmale den Agenten am stärksten in seinen Aktionen (z.B. Sammelschienenwechsel) beeinflussen. Darüber hinaus ermöglicht SHAP die Bewertung, weshalb der Agent für eine ausgewählte Netzsituation genau diesen Schaltzustand vorschlägt.

Ergänzend werden Entscheidungsbäume (*Decision Trees*) angewandt, um interpretierbare Regeln abzuleiten. Ein Entscheidungsbau ist ein Verfahren, das Daten durch eine hierarchische Abfolge von binären Entscheidungen klassifiziert. Ausgehend von einem Wurzelknoten wird der Datensatz sukzessive anhand von Schwellenwerten einzelner Merkmale (*Features*) in Teilmengen aufgespalten, bis die Blattknoten eine eindeutige Klassenzuordnung ermöglichen. Die Auswahl der Splitkriterien erfolgt durch Metriken wie den Gini-Index oder die Informationsentropie, die den Informationsgewinn maximieren [3]. Im vorliegenden Anwendungsfall approximiert der Entscheidungsbau das Verhalten des trainierten Agenten und extrahiert daraus lesbare If-Then-Regeln, die dem Betriebsplaner oder Operator die Bewertung der Topologievorschläge erleichtern.

## Erkenntnisse

Der Beitrag zeigt eine Topologieoptimierung eines Übertragungsnetzes für einen beispielhaften Dezembertag – die Topologieoptionen sind mit einer Netzsicherheitsrechnung überprüft und mit SHAP untersucht. Die SHAP-Werte lassen sich nicht vollständig ermitteln, da die notwendigen *Logits* bzw. *Probabilities* nicht zielführend extrahiert werden können. Dies macht Schwachstellen im Modell sichtbar, die u.a. mit einer Vergrößerung des Neuronalen Netzes sowie einer entsprechenden Hyperparameteroptimierung adressiert werden.

Ergänzend ermöglichen die aus *Decision Trees* abgeleiteten Regeln eine zügige Bewertung der Topologieoptimierung. Es zeigt sich, dass der Agent einen vorwiegend zeitlichen Verlauf erlernt; die Schaltzustände werden stark von Jahres- und Tageszeit beeinflusst. Für Querkupplung 4431 ist beispielsweise festzuhalten, dass diese an Wintertagen in der Spätschicht (ab ca. 18:30 Uhr) sowie Nachschicht stets geöffnet und in der Frühschicht eher geschlossen betrieben wird – was mit dem operativen Tagesgeschäft einer Ausschaltplanung harmoniert. Als If-Then-Regel für den Operator formuliert: Der Schalter folgt der elektrischen Last – er öffnet in den Abend-, Nacht- und Morgenstunden (sinkende Last) und schließt tagsüber (steigende Last).

## Referenzen

- [1] Antoine Marot, Benjamin Donnot et. al. Learning to run a Power Network Challenge: a Retrospective Analysis. On *arXiv:2103.03104*, 2021.
- [2] Daniel Beechey, Thomas M.S. Smith und Özgür Simsek. Explaining Reinforcement Learning with Shapley Values. In *Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii*, 2023.
- [3] Kevin P. Murphy. Probabilistic Machine Learning: An Introduction. In *The MIT Press, Cambridge, Massachusetts*, 2022.