

Weißes Rauschen basierte Verlustfunktion zur verbesserten Zeitreihenprognose mit künstlichen neuronalen Netzen

Tobias Fischer, Fabian Bauer, Silas Selzer, Peter Bretschneider

Technische Universität Ilmenau, Institut für Elektrische Energie- und
Steuerungstechnik, Fachgebiet Energieeinsatzoptimierung
Gustav-Kirchhoff-Str. 5, 98693 Ilmenau
E-Mail: {tobias-merlin.fischer, fabian.bauer, silas-aaron.selzer,
peter.bretschneider}@tu-ilmenau.de

1 Einleitung

Zeitreihenprognosen stellen ein wichtiges Werkzeug für eine sichere und stabile Energieversorgung dar. Exakte Vorhersagen von Energielastgängen bilden die Grundlage für eine optimale Planung und Steuerung des elektrischen Energiesystems. Zur möglichst genauen Vorhersage ist es notwendig, Modelle zu entwickeln und zu untersuchen, die dem wachsenden Mengengerüst an messtechnisch erfassten Energiedaten und exogenen Einflussgrößen Rechnung tragen. Vor diesem Hintergrund spielen Verfahren des maschinellen Lernens eine wichtige Rolle bei der Modellierung linearer und nicht-linearer Zeitreihen. Die Qualität und Quantität der Informationsgewinnung wird beim maschinellen Lernen über die verwendete Verlustfunktion bewertet. In den durchgeführten Untersuchungen werden drei überarbeitete Verlustfunktionen basierend auf weißem Rauschen entwickelt und mit bisherigen Methoden verglichen.[1, 2, 3]

2 Weißes Rauschen und Periodogramm

Der Einsatz des Gradientenabstiegsverfahrens im maschinellen Lernen garantiert nur die Konvergenz zu einem stationären Punkt, was sich als problematisch erweisen kann. Typische Verlustfunktionen, wie bspw. der mittlere quadratische Fehler (MSE), stellen ein Optimalitätskriterium basierend auf dem Amplitudenabstand der Vorhersage und den Originaldaten dar. Dabei ist zu bedenken, dass es Ziel der Regressionsanalyse ist, Muster in einer Zeitreihe zu erlernen, der MSE aber nur deren Ausprägung als Zielwert definiert. Um die Zeitachse als Zieldimension mitzubetrachten, wird der MSE um die Untersuchung des Residuums, also der Differenz aus Prognose und Originaldaten, auf weißes Rauschen erweitert. Eine Vorhersage kann als optimal angenommen werden, wenn das Residuum weißem Rauschen entspricht und somit das Restsignal keine Informationen enthält. In dieser Arbeit wird weißes Rauschen mit Gauß'schen weißen Rauschen gleichgesetzt. Gaußsches weißes Rauschen ist ein Prozess, dessen Zufallsvariablen ε_t normalverteilt sind und alle den gleichen Erwartungswert $\mu = E(\varepsilon_t)$ und Varianz $\sigma^2 = \text{Var}(\varepsilon_t)$ haben. Er ist der einfachste schwachstationäre Prozess, Mittelwert- und Varianzfunktion sind hier für alle t konstant $\mu(t) = \mu$, $\sigma^2(t) = \sigma$. Der Vorteil des weißen Rauschens gegenüber anderem Rauschen, bspw. dem $1/f$ -Rauschen ist die stochastische Unabhängigkeit der Zufallsvariablen. Damit gilt, für $s \neq t$, $E(\varepsilon_s \varepsilon_t) = E(\varepsilon_s)E(\varepsilon_t)$, voraus $\text{Cov}(\varepsilon_s, \varepsilon_t) = 0$ folgt. Das Spektrum des weißen Rauschens ist konstant, alle Frequenzen liefern den gleichen Beitrag zur Prozessvarianz. Dieses konstante Spektrum ermöglicht es mittels geeigneter Leistungsdichteschätzer einen Test auf weißes Rauschen durchzuführen. Eine Teststatistik zur Überprüfung der Hypothese, ob der untersuchte Prozess weißem Rauschen entspricht, ist das sogenannte kumulierte Periodogramm. Für den genauen Ablauf der Teststatistik wird auf Unterabschnitt 3.2 verwiesen. [4, 5, 6]

3 Untersuchungsansätze

Die Prognose der Energielastgänge erfolgt mittels künstlicher neuronaler Netze und als Netzarchitektur wird ein Multi-Layer-Perceptrons (MLP) verwendet.

Die Vorhersage der gesuchten Größe zum nächsten Zeitschritt $Y(t_n)$ erfolgt aus den letzten N Werten der Eingangsvariablen $(X(t_{n-1}) \dots X(t_{n-N}))$. Mit $X \subseteq \mathbb{R}^{d_x}$ als Eingaberaum und $Y \subseteq \mathbb{R}^{d_y}$ als Ausgaberaum ergibt sich als Ziel des überwachten Lernens eine zu minimierende Verlustfunktion $L((x; \theta); y)$, wobei θ sowohl die Parameter (Gewichte), als auch Hyperparameter bezeichnet.

3.1 Mean Squared Error

Eine der am häufigsten eingesetzten Verlustfunktionen in der Regression ist der Mean Squared Error, aufgrund der erhöhten Bestrafung von stärkeren Abweichungen und wird berechnet durch [7]:

$$MSE = \frac{1}{N} \sum_{n=1}^N \left(x_{n_{\text{pred}}} - y_{n_{\text{true}}} \right)^2 \quad (1)$$

3.2 Weißes Rauschen basierte Verlustfunktionen

Der hier verwendete Ansatz erweitert den MSE um einen Summanden, der den Fehler zwischen Vorhersage und Realwerten nicht nur als Distanzdifferenz berechnet, sondern das Residuum $R_n = x_{n_{\text{pred}}} - y_{n_{\text{pred}}}$ mit $n = 1, \dots, N$ auf weißes Rauschen untersucht. Der Einflussfaktor α entspricht dem relativen Einfluss der erweiterten Verlustfunktion auf den MSE. β ist ein Skalierungsfaktor, der Größenunterschiede zwischen beiden Funktionen ausgleicht. Es werden drei Möglichkeiten untersucht, die Gleichheit des Residuums und weißem Rauschen für das KNN zu bestimmen. Indem ein Signal W als ein Signal Gauß'schen weißen Rauschens definiert wird, mit Standardabweichung und Mittelwert identisch zu R , kann mit der Fourierfrequenz $\lambda_n = n/N$ und Autokovarianzfunktion c_τ für $r = 1, \dots, N$ die Verlustfunktion mittels Fourier Transformation erweitert werden. [5, 8]

3.3 Kumuliertes Periodogramm

Das kumulierte Periodogramm (KP), dargestellt durch

$$KP_r = \frac{\sum_{n=1}^r \sum_{\tau=-(N-1)}^{N-1} c_\tau \cdot \exp(2\pi i \lambda_n \tau)}{\sum_{n=1}^N \sum_{\tau=-(N-1)}^{N-1} c_\tau \cdot \exp(2\pi i \lambda_n \tau)} \quad (2)$$

als Teststatistik der Hypothese des weißen Rauschen basiert auf den Überlegungen in Abschnitt 2 und die Berechnung ist entlehnt an [5]. Da das Spektrum des weißen Rauschens konstant ist und KP_r den Anteil spektraler Masse der ersten r Fourierfrequenzen angibt, sollte das kumulierte Periodogramm um die Winkelhalbierende des Einheitsquadrates schwanken, wenn die Hypothese zutrifft. [5] Die verwendete Verlustfunktion erweitert sich somit zu:

$$L_{\text{Periodogramm}} = \alpha \cdot \beta \cdot \frac{1}{N} \sum_{r=1}^N KP_r + (1 - \alpha) \cdot MSE \quad (3)$$

Wichtig ist dabei zu erwähnen, dass das Periodogramm zwar ein Erwartungstreuer, aber kein konsistenter Schätzer ist. Dies bedeutet, dass die Varianz des Periodogramms für eine steigende Anzahl betrachteter Zufallsvariablen nicht gegen Null konvergiert. In der Literatur werden mehrere Vorschläge unterbreitet dem entgegenzuwirken. Unter anderem wird in dieser Arbeit der Ansatz des Moving Average Filters von Bartlett verwendet, welcher für die Prognose die Walking Forward Validierung nutzt. [4, 6]

3.4 White Noise Based Loss

Eine Erweiterung des kumulierten Periodogramms stellt der White Noise Based Loss (WNBL) mittels

$$L_{\text{WNBL}} = \alpha \cdot \beta \cdot \frac{1}{N} \sum_{r=1}^N (KP_r(R) - KP_r(W)) + (1 - \alpha) \cdot MSE \quad (4)$$

dar, der die Zielrichtung der Verlustfunktion durch die Abweichung des kumulierten Periodogramms der Restsequenz $KP(R)$ von einem kumulierten Periodogramm weißen Rauschens $KP(W)$ explizit vorgibt.

3.5 White Noise Signal Loss

Der White Noise Signal Loss (WNSL) ist die dritte vorgeschlagene Verlustfunktion, um die Gleichheit von R und weißem Rauschen festzustellen und definiert diese auf Signalebene. Die Gleichung

$$L_{\text{WNSL}} = \alpha \cdot \beta \cdot \frac{1}{N} \sum_{n=1}^N (r_n - w_n)^2 + (1 - \alpha) \cdot \text{MSE} \quad (5)$$

optimiert den punktweisen quadrierten Fehler der Restsequenz R und dem Signal W .

4 Modellierung

Die Datengrundlage besteht aus einer plausiblen und vollständigen Zeitreihen für 3 Jahre mit einer zeitlichen Auflösung von 15 Minuten. Für alle Untersuchungsszenarien wird der Week-Ahead Prognosehorizont festgelegt, bei einer Eingangssequenzlänge von einem Monat, was einem Verhältnis von 2880 auf 672 Sequenzpunkten entspricht. Als Modell wird ein MLP verwendet mit 2 verdeckten Schichten mit jeweils 24 Neuronen. Die Aktivierungsfunktion ist ReLU und der eingesetzte Optimierer Adam. [5] Das MLP wurde ausgewählt, da es im Bereich des Deep Learning die Standard Netz- und Neuronenarchitektur darstellt und die Ergebnisse so besser in den bestehenden Forschungsstand eingeordnet werden können.

5 Simulative Untersuchung und Evaluierung

Ein Auszug der Ergebnisse der Untersuchung sind in Tabelle 1 dargestellt. Die Verlustfunktion des White Noise Based Loss und kumulierten Periodogramm

Tabelle 1: Methode der Verlustfunktion und Fehlermaß

Verlustfunktion	MSE	Periodogramm	WNSL	WNBL
RMSE	0.968	0.721	0.957	0.721
MAPE (%)	9.66	7.49	9.43	7.48
MAE	0.737	0.572	0.72	0.571
α	0	0.99	0.1	0.99

erzielte in allen gemessenen Fehlermaßen die besten Ergebnisse. Eine verbesserte Generalisierungsfähigkeit kann bei Untersuchung der Prognosegüte über die gesamte Sequenzangabe festgestellt werden. Abbildung 1 zeigt den RMSE der MSE- und WNBL-Verlustfunktion, deutlich zu erkennen ist ein positiver Trend der MSE-Verlustfunktion über die Sequenzlänge, ebenso wie die Minima an den Autokorrelationsspitzen im Datensatz, den Tagesrhythmen (96 Sequenzpunkte). Im Vergleich dazu nähert sich der Verlustverlauf des WNBL mehr dem Idealverlauf eines vollständig unabhängig zufälligen Messfehlerverlaufs (siehe Abbildung 1).

Der durchgeführte Vergleich von Methodiken für Verlustfunktionen für künstlicher neuronaler Netze zur Energielastprognose hat gezeigt, dass durch die Verwendung von erweiterten Verlustfunktionen auf Basis des weißen Rauschens die Prognosegüte und Konvergenz gesteigert werden kann.

6 Zusammenfassung und Ausblick

Im Ergebnis der Untersuchung konnte bestätigt werden, dass die verwendeten Varianten der Verlustfunktionen einen nachweisbaren positiven Effekt auf die Modellperformance hinsichtlich der Generalisierungsfähigkeit und Robustheit haben. Von allen drei untersuchten Verlustfunktionen, basierend auf weißem Rauschen, erzielte der White Noise Based Loss die besten Ergebnisse. Im direkten Vergleich weißes Rauschen basierter Verlustfunktionen zu bereits etablierten, wie dem MSE, ist ein erhöhter Justierungsaufwand durch den Einfluss- und Skalierungsfaktor gegeben, was dem Model einen zusätzlichen optimierbaren Hyperparameter hinzufügt

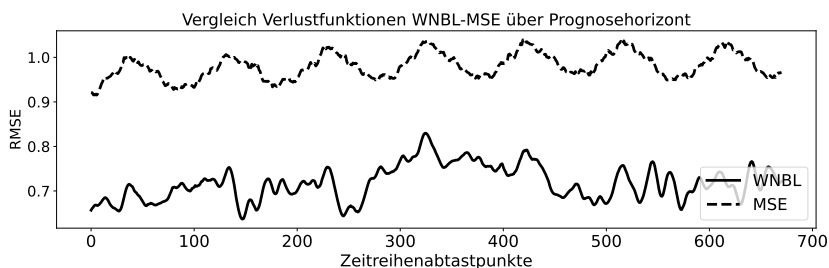


Bild 1: RMSE der MSE- und WNBL-Verlustfunktion

Literatur

- [1] I. Goodfellow, Y. Bengio und A. Courville. „Deep Learning: Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze“. Frechen: mtip. 1. Auflage. 2018.
- [2] L. Fahrmeir, T. Kneib und S. Lang. „Regression: Modelle, Methoden und Anwendungen“. Berlin, Heidelberg: Springer Berlin Heidelberg. 2. Auflage 2009.
- [3] O. D. Doleski. „Herausforderung Utility 4.0: Wie sich die Energiewirtschaft im Zeitalter der Digitalisierung verändert“. Wiesbaden: Springer Vieweg. 2017.
- [4] J.-P. Kreiß und G. Neuhaus. „Einführung in die Zeitreihenanalyse“. Berlin, Heidelberg: Springer Berlin Heidelberg. 2006.
- [5] R. Schlittgen und B. H. J. Streitberg. „Zeitreihenanalyse“. München, Oldenbourg: Oldenbourg. 9. Auflage. 2011.
- [6] L. H. Koopmans. „The Spectral Analysis of Time Series“. Erscheinungsort nicht ermittelbar: Elsevier Science & Technology. 2. Auflage. 2011.
- [7] Z. Wang und A. C. Bovik. „Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures“. In: IEEE Signal Process. Mag., Jg. 26, Nr. 1, S. 98-117. 2009.

- [8] X. Cui, W. Zhang, Z. Tüske und M. Picheny. „Evolutionary Stochastic Gradient Descent for Optimization of Deep Neural Networks“. arXiv-preprint. 1810.06773v1. 2018.
- [9] D. P. Kingma und J. Ba. „Adam: A Method for Stochastic Optimization“. arXiv-preprint. arXiv:1412.6980v9. 2014.