Windertragsprognose mit Echo State Networks

Dr. Hans-Ulrich Kobialka, Fraunhofer IAIS, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
Hans-Ulrich.Kobialka@iais.fraunhofer.de

Kurzfassung

Statistische Verfahren können komplexe dynamische Zusammenhänge erfassen, die es nicht nur erlauben, nicht-observierbare Größen zu schätzen, sondern auch zukünftige Größen zu prognostizieren. In diesem Papier wird dies an Hand der Erstellung einer Windertragsprognose beschrieben. Hierzu wird das Verfahren „Echo State Networks“ verwendet, das es gestattet, sehr große und leistungsfähige rekurrente neuronale Netzwerke zu trainieren. Dieses Papier gibt einen Überblick zu dem Echo-State-Verfahren und beschreibt die technische Realisierung der erstellten Windertragsprognose.

Abstract

Statistical methods are able to create models of complex system dynamics which are difficult to capture analytically. This paper describes a wind energy prediction system based on a machine learning method, called Echo State Networks. Echo State Networks enable the training of large recurrent neural networks which are able to model and predict highly non-linear system dynamics. This paper gives a short description of Echo State Networks and the realization of the wind energy prediction system.

1 Einleitung

Die Prognose des Energie-Ertrags von Sonnen- und Windenergie ist von großer Wichtigkeit, um im Mangelfall Ausgleichs-Energie rechtzeitig zu allokieren bzw., im Falle eines Überschusses, diesen potentiellen Verbrauchern zu signalisieren. Je früher und präziser ein Energie-Ertrag prognostiziert werden kann, desto flexibler und damit kostenoptimiert können Marktteilnehmer sich darauf einstellen.

Auch für automatisierte Steuerungs- und Regel-Prozesse (z.B. zur automatischen Lastverteilung in Stromnetzen) können Prognose-Verfahren und deren Güte von großer Wichtigkeit sein („Prädiktive Regelung“).

Im Folgenden wird die Realisierung einer Windertragsprognose beschrieben, die durch Fraunhofer IAIS im Auftrag der Firma Amprion für deren Regelzone erstellt wurde.

Die maßgebenden Einflussgrößen für eine Windertragsprognose sind Wetterprogosen und Informationen über den Zustand der Windturbinen und Windparks (aktuelle Ertragsleistung, Verfügbarkeit, geplante Wartungsarbeiten, etc.). Im konkreten Fall waren die Einflussgrößen die aktuelle Ertragsleistung von 22 Windparks in der Regelzone Amprion und Wetterprognosen für die Standorte dieser 22 Windparks. Diese Standorte waren topologisch mehr oder weniger ungleichmäßig verteilt. Außerhalb dieser 22 Windparks gibt es in der Regelzone Amprion noch über 2000 Windenergie-Anlagen (WEA), die mal nahe, mal weit entfernt von den 22 beobachteten Windparks gelegen sind.

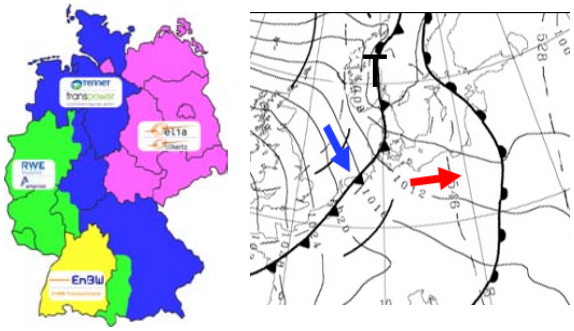

Abbildung 1: Regelzone Amprion (links, grün) und ein Beispiel für durchziehende Wetterfronten (rechts).
Bildquellen: Wikipedia, UKMO http://profi.wetteronline.de

Die je nach Windrichtung und -stärke verschiedenen kausalen und zeitlichen Korrelationen zwischen den 22 Windparks und den übrigen sind (auch bedingt durch Topologie und Thermodynamik) höchst komplex und vielfältig.

Beispielsweise hierfür wäre der Zusammenhang, wie eine an einem Standort gemessene Windgeschwindigkeit abhängig von Windrichtung, Luftdruckverteilungen, Temperatur etc. sich später auf die Windenergie-Leistung eines 50 km entfernten Windparks (dessen Windertragsleistung zudem unbekannt ist) auswirkt. Hierfür topologische und thermodynamische Simulations-Modelle zu erstellen und aus ihnen Prognosen zu berechnen, ist sehr aufwendig und rechenintensiv. Natürlich gibt es in der Regelzone Amprion sehr viele solcher Zusammenhänge.

Daher wurde von einer physikalischen Modellbildung Abstand genommen und stattdessen ein statistisches Lernverfahren eingesetzt.

2 Statistische Lernverfahren

Statistische Lernverfahren lernen Gesetzmäßigkeiten eines Systems aus Daten. Statistisch gesehen sind Gesetzmäßigkeiten Korrelationen.

Da sich beobachtete Ereignisse (z.B. eine durchziehende Wetterfront) sich an verschiedenen Standorten zeitlich verzögert auswirken, kommen nur Verfahren zur Zeitreihen-Analyse und -Prognose in Frage. Dazu gehören nichtlineare auto-regressive Verfahren (NARX), Volterra Reihen und rekurrente neuronale Netzwerke (RNNs), z.B. time delay RNNs [1] [2] [7].

„Echo State Networks“ (ESNs) [3] [4] [6] ist ein von Fraunhofer IAIS patentiertes Lernverfahren für RNNs. RNNs (und damit auch ESNs) haben im Gegensatz zu sog. Feedforward Neuronalen Netzen (FNNs) Rückkopplungsschleifen. Ein RNN stellt mathematisch ein dynamisches System (vergleichbar einer Menge von Differenzialgleichungen) und somit eine Funktion auf der Zeitreihe der eingelesenen Messwerte dar.

Der Vorteil der ESN-Technologie gegenüber anderen Verfahren zum Training von RNNs liegt darin, dass das Training nicht-iterativ, und damit schneller und stabiler erfolgt. Dies ermöglicht es, sehr große Netze (mehrere tausend Knoten) zu verwenden, die die Fähigkeit haben, das Verhalten von sehr komplexen Systemen zu lernen und zu prognostizieren.

Im vorliegenden Fall musste ein großes Netz (mindestens 500 interne Knoten) verwendet werden, allein schon wegen der großen Zahl der Eingangsgrößen (176) und der aufgrund verschiedener Wettersituationen großen Zahl von Wechselwirkungen zwischen ihnen (wie oben beschrieben).

3 Echo State Networks

In diesem Abschnitt wird ein an einigen Stellen leicht vereinfachter Überblick über das Verfahren der Echo State Networks gegeben.

Ein Echo State Network ist ein rekurrentes neuronales Netz, welches zum Zeitpunkt t ein Ergebnis $y(t)$ berechnet basierend auf seinem aktuellen internen Zustand $x(t)$.

$$y(t) = f_{out}\,(x(t)^T\, W_{out})\ . \qquad (1)$$

Hierbei bezeichnet W_{out} die Matrix der Ausgangsgewichte und ^{T} ist die Notation für die Transposition einer Matrix. $f_{out}(.)$ ist die sogenannte Output-Aktivierungsfunktion. In der folgenden Beschreibung verwenden wir der Einfachheit halber hierzu die Identitätsfunktion ($f_{out}(x) = x$).

Der Zustand $x(t)$ der internen Knoten wird aus den aktuellen Eingangswerten $u(t)$ und dem vorhergehenden Zustand $x(t-1)$ berechnet. Zusätzlich kann auch die letzte Ausgabe der Netzes $y(t-1)$ verwendet werden.

$$x(t) = f\,(W_{in}\, u(t) + W\, x(t-1) + W_{fb}\, y(t-1))\ . \qquad (2)$$

W, W_{in}, und W_{fb} sind die Gewichtsmatrizen der Verbindungen zwischen den internen Knoten (W), zwischen den Eingangs-Knoten und dem Netz (W_{in}) und zwischen den Ausgangs-Knoten und dem Netz (W_{fb}).

Ein ESN wird wie folgt erzeugt. Entsprechend der zu lösenden Aufgabe wird die Zahl der Eingangs- und Ausgangs-Knoten bestimmt. Die Anzahl der internen Knoten kann frei gewählt werden, entsprechend der benötigten Lernkapazität. Im einfachsten Fall können die Gewichte der Matrizen W, W_{in} und W_{fb} zufällig zwischen 0 und 1 generiert werden. Die Matrix W für die Verbindungen zwischen den internen Knoten muss dünn besetzt sein, d.h. nur ein kleiner Anteil der Gewichte (z.B. 10%) darf ungleich Null sein. Matrix W muss zudem die sogenannte Echo State Eigenschaft aufweisen [5]. In vielen (aber nicht in allen) Fällen genügt es, diese Matrix so zu skalieren, dass der spektrale Radius kleiner als Eins ist. Einzelheiten und Hintergründe werden in [8] erläutert.

Nachdem ein ESN erzeugt worden ist, werden die Trainingsdaten (bestehend aus den Eingangsdaten u_{train} und den dazugehörigen Ausgangsdaten y_{train}) sukzessive in das Netz gefüttert. Dabei werden die Zustände des Netzes $x_{train}(t)$ immer wieder, analog zu Gleichung (2), berechnet.

$$x_{train}(t) = \qquad (3)$$
$$f\,(W_{in}\, u_{train}(t) + W\, x_{train}(t-1) + W_{fb}\, y_{train}(t-1))$$

Alle nacheinander berechneten Trainingszustände $x_{train}(t)$ werden in der Matrix X_{train} gesammelt. Ebenso werden alle Ausgabewerte y_{train} in der Matrix Y_{train} zusammengefasst. Falls nur ein Ausgabewert berechnet werden soll, ist Y_{train} ein Vector.

$$X_{train} = \begin{bmatrix} x_{train}^T(1) \\ x_{train}^T(2) \\ \vdots \\ x_{train}^T(N_{train}) \end{bmatrix}\ ,\ Y_{train} = \begin{bmatrix} y_{train}(1) \\ y_{train}(2) \\ \vdots \\ y_{train}(N_{train}) \end{bmatrix} \qquad (4)$$

N_{train} ist die Länge der Trainingsdaten. Entsprechend Gleichung (1) und mit $f_{out}(x) = x$, muss nun folgendes gelten.

$$Y_{train} = X_{train}\, W_{out}. \qquad (5)$$

Da X_{train} and Y_{train} gegeben sind, kann die Matrix der Ausgangsgewichte W_{out} mittels Regression berechnet werden. Dies erfolgt für jeden Ausgangsknoten separat.

Es können wahlweise verschiedene Regressions-Verfahren eingesetzt werden. Ein einfaches aber effektives Verfahren ist die sogenannte Ridge-Regression (auch bekannt als Tikhonov Regularization). Hier wird ein Regularisierungs-Parameter λ dazu benutzt, hohe Gewichts-Werte zu vermeiden und damit die Empfindlichkeit gegenüber dem Rauschen von Eingangswerten zu reduzieren. Damit kann eine Überanpassung (englisch ‚Overfitting') vermieden werden.

$$W_{out} = (X_{train}^T X_{train} + \lambda^2 I)^{-1} X_{train}^T Y_{train} \ . \qquad (6)$$

wobei I die Identitäts-Matrix darstellt und $^{-1}$ die Invertierung einer Matrix.

Auch nicht-lineare Regressionsverfahren (z.B. SVM-Regression oder Gaussian Process Regression) können angewandt werden.

4 Technische Realisierung der Windertragsprognose

Input für eine Windertragsprognose sind 176 Zeitreihen: Windertragsmessungen an 22 Standorten, und 22 Wetterprognosen mit je 7 Wetterinformationen. Die Verwendung von Echo State Networks verlangt nicht die Zahl der Eingangsgrößen einzuschränken, oder deren Relevanz für die Prognose abzuschätzen. Sogenannte Feature-Selection-Verfahren, die versuchen die beste Kombination von Eingangsgrößen zu ermitteln, wären bei einer großen Zahl von Eingangsgrößen (wie dies hier der Fall ist) sehr aufwendig.

Alle Zeitreihen beginnen mit den Werten, die 10 Stunden vom aktuellen Zeitpunkt zurückliegen. Die Zeitreihen der Messungen enden mit der jeweils letzten verfügbaren Messung. Die aus Wetterprognosen extrahierten Zeitreihen enden zu dem Zeitpunkt, für den eine Ertragsprognose erstellt werden soll (**Abbildung 2**).

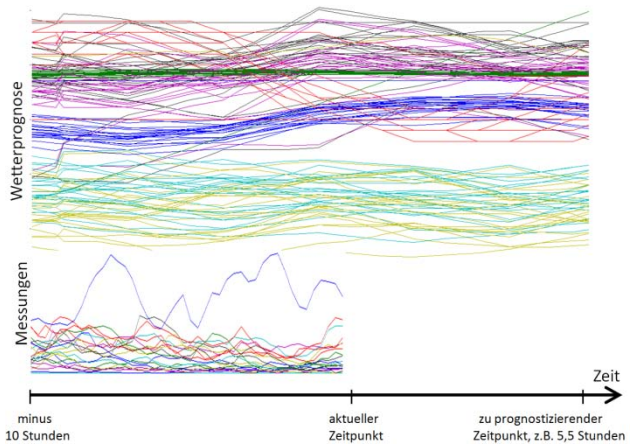

Abbildung 2: Zeitreihen, die für eine Windertragsprognose (z.B. für 5,5 Stunden in der Zukunft) verwendet werden.

Das Ergebnis der Windertragsprognose ist der Windertrag (gemessen in Megawatt) für die gesamte Regelzone Amprion für 33 Zeitpunkte in der Zukunft, angefangen mit der Prognose für die aktuelle Viertelstunde, und weiter (in Viertelstundenschritten) bis hin zu der Prognose des Windertrags in acht Stunden.

Für jeden der 33 zu prognostizierenden Viertelstundenwerte wurde eine separate Prognose trainiert. Die 154 Wetterprognose-Zeitreihen und die 22 Messdaten-Zeitreihen werden zunächst getrennt durch ein ESN in je einen hochdimensionalen Zustandsraum transformiert. Anschließend werden beide Zustände linear fusioniert und der gewünschte Prognosewert berechnet.

Als Trainingsdaten wurden Messungen und Wetterprognosen von zweieinhalb Jahren (Anfang 2009 bis Mitte 2011) verwendet. In diesem Zeitraum wurden 21.681 Zeitpunkte ermittelt, für die dann jeweils eine 154-dimensionale Wetter-Zeitreihe und eine 22-dimensionale Messwert-Zeitreihe erzeugt wurden (Abbildung 2).

Diese Daten enthielten fehlende bzw. ungültige Werte. Sofern diese Datenlücken nur ein oder zwei Werte umfassten, wurden sie zeitlich interpoliert. Falls eine Zeitreihe größere Lücken enthielt, wurde der entsprechende Zeitpunkt verworfen. Hiernach blieben 19.101 Zeitpunkte.

Zur Optimierung einiger ESN-Parameter wurden diese Daten in Trainings- und Test-Datenmengen aufgeteilt und diverse Tests durchgeführt. Abschließend wurde mit allen 19.101 Zeitpunkten (insgesamt 322.730.496 Zahlenwerte) das finale Windertragsprognose-Modell trainiert.

Die ESNs werden am Fraunhofer IAIS trainiert und anschließend als Datei in die EDV-Umgebung der Firma Amprion transferiert. Dort liest ein in C++ geschriebenes Ausführungs-Modul bei Start ein ESN ein. Im laufenden Betrieb liest dieses Modul kontinuierlich eintreffende Messdaten und Wetter-Prognosen ein und erstellt viertelstündlich bzw. auf Knopfdruck die Prognosen für die nächsten 8 Stunden.

Diese Realisierung kommt ohne Kommunikationsverbindung zum Fraunhofer IAIS aus. Im Falle von fehlenden Messdaten (wegen Ausfall einer Messstation oder der Kommunikation dorthin) werden die letzen verfügbaren Werte als weiterhin gültig angenommen und eine Warnung an das Bedienpersonal gesendet. Falls jedoch Messdaten über einen längeren Zeitraum nicht zur Verfügung stehen, oder neue Wetterprognosen nicht wie vereinbart eintreffen, kann dies nicht mehr kompensiert werden und die Ergebnisse der Windertrags-Prognose sind nicht mehr valide.

Es wird ein ESN mit 1000 internen Knoten verwendet. Das Erstellen der 33 Prognosen für die nächsten 8 Stunden (in Viertelstundenschritten) dauert ca. zwei Minuten.

5 Ergebnisse

Das hier beschriebene Windertragsprognose-System wird seit Anfang November 2011 bei der Firma Amprion im täglichen Betrieb eingesetzt. Die Abweichungen der Prognose von den später tatsächlich erzielten Winderträgen wurden von Amprion ermittelt und sind beispielhaft in **Abbildung 3** dargestellt.

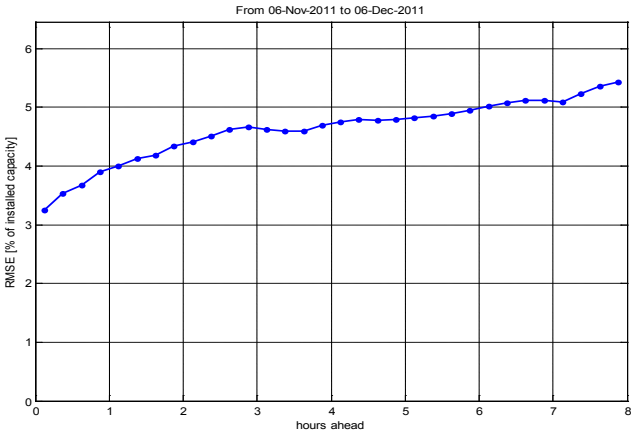

Abbildung 3: RMSE gemittelt über den Zeitraum 06.11. bis 06.12.2011 für verschiedene Prognose-Zeiträume (0,25 bis 8 Stunden). Graphik: Amprion

Die Abweichungen werden als mittlerer quadratischer Vorhersagefehler (RMSE) gemittelt über den Zeitraum 06.11.2011 bis 06.12.2011 berechnet. Maßeinheit ist Prozent der installierten Windenergiekapazität in der Regelzone Amprion. Diese Berechnung erfolgte getrennt für jeden der hier dargestellten 32 Prognose-Zeiträume (0,25 bis 8 Stunden in die Zukunft).

6 Ausblick: Prognosen zur Optimierung in Smart Grids

Die Vision eines Smart Grids beinhaltet a) die wachsende Rolle von regenerativen Energien und b) die dynamische Interaktion von Strom-Erzeugern und –Verbrauchern. In diesem Umfeld sind Prognosen von hoher Wichtigkeit, nicht nur um mit naturbedingten Energieschwankungen besser umgehen zu können, sondern ebenso um das Verhalten von Marktteilnehmern und technischen Systemen besser kalkulierbar zu machen.

Die in Prognosen verwendbaren Einflussfaktoren sind oft sehr unterschiedlich („Äpfel und Birnen“) bzw. enthalten indirekt gemessene Größen, deren Zusammenhang zu der zu prognostizierenden Größe sich nur unzureichend formal beschreiben lässt. Beispielsweise wird nur punktuell gemessen und nicht eine gesamte Fläche abgedeckt. Oder eine gemessene Größe wirkt sich je nach Situation (z.B. Windgeschwindigkeit) verzögert auf die zu prognostizierende Größe aus. Oder es wird mit Messgeräten gemessen, die bestimmte Eigenschaften haben (z.B. Maximalwerte, Rauschverhalten) und sich individuell unterscheiden. Solche Phänomene lassen sich statistisch gut erfassen.

Das Verhalten der überwiegenden Zahl der zu prognostizierende Systeme ist zudem zeitabhängig, d.h. abhängig von vorangegangenen Ereignissen, die den Zustand eines Systems und damit auch sein zukünftiges Verhalten beeinflussen. Wie diverse Ereignisse einen in der Regel nicht vollständig observierbaren Zustand eines Systems beeinflussen, lässt sich im Grunde eigentlich nur statistisch feststellen. Daher werden statistische Verfahren zur Zeitreihen-Analyse und -Prognose innerhalb von Smart Grids eine wichtige Rolle spielen.

7 Literatur

[1] Giannakis G.B & Serpendin E: *A bibliography on nonlinear system identification.* Signal Processing, 81 2001 533-580

[2] Haykin S.: *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1999.

[3] Jaeger, H., Haas, H.: Harnessing Nonlinearity: *Predicting Chaotic Systems and Saving Energy in Wireless Communication*. Science, April 2, 2004:78-80.

[4] Jaeger, H., Maass, W., Principe J. (eds): *Special Issue Echo State Networks and Liquid State Machines*, Neural Networks Volume 20, Issue 3, pp 287-432, 2007.

[5] Jaeger H.: *The "echo state" approach to analysing and training recurrent neural networks*. GMD Report 148, German National Research Center for Information Technology, 2001.

[6] Lukosevicius, M., Jaeger, H.: *Reservoir computing approaches to recurrent neural network training*. Computer Science Review, 3(3):127-149, 2009.

[7] Nelles, O: *Nonlinear System Identification*. Springer Berlin, ISBN 3-540-67369-5, 2000.

[8] Yildiz, I. B., Jaeger, H, Kiebel, S. J.: *Re-Visiting the Echo State Property*. to appear in Neural Networks, 2012.