

Deep Learning zur Vorhersage von Feinstaubbelastung

Georges Alkhouri,¹ Moritz Wilke²

1 Einleitung

Feinstaubbelastung steht seit einiger Zeit in der öffentlichen Debatte und stellt mir hoher Wahrscheinlichkeit ein großes Gesundheitsrisiko dar. Laut WHO [Or06] kann die Reduzierung von Feinstaub zur Senkung verschiedener Krankheiten, wie bspw. Herzinfarkten, Lungenkrebs und asmathischen Erkrankungen dienen. Deswegen werden von der Organisation Tagesgrenzwerte von $25 \mu\text{g}/\text{m}^3$ für Partikel um $2,5 \mu\text{m}$ (PM_{2,5}) und $50 \mu\text{g}/\text{m}^3$ für Partikel um $10 \mu\text{m}$ (PM₁₀) empfohlen. In diesem Beitrag zur Data Science Challenge soll gezeigt werden, wie die vorhandenen Feinstaubsensoren in der Stadt Leipzig genutzt werden können, um zukünftige Werte vorherzusagen.³ Eine solche Vorhersage könnte nicht nur zur Warnung dienen, sondern auch Grundlage für kurzfristige Gegenmaßnahmen (bspw. den Wechsel auf ÖPNV) bilden.

2 Vorverarbeitung der Sensordaten

Die vorhandenen Daten in Leipzig bringen zwei Herausforderungen mit sich. Zum Einen braucht es für eine zuverlässige Verwendung der Feinstaubmesswerte aus einem Sensor vom Typ SDS011 auch ein Luftfeuchtigkeitswert zum gleichen Zeitpunkt, da die Sensoren nur zwischen einer relativen Luftfeuchtigkeit von 20% – 50% zuverlässig messen. [Bu18] Diese Messungen sind nicht für jeden Ort vorhanden. Zum Anderen benötigt es für eine Vorhersage eine Messreihe von mindesten einem Jahr, wenn (vermutete) zyklische Einflussfaktoren, wie Jahreszeit, berücksichtigt werden sollen.

Abb. 1 zeigt alle Standorte mit Sensoren vom Typ SDS011 und DHT22 (Temperatur, Luftfeuchtigkeit) in Leipzig. Da deren Verteilung keine genauere örtliche Unterteilung innerhalb der Stadt ermöglicht und die Sensoren z.T. erst seit Anfang 2018 in Betrieb

¹ Competence Center for Scalable Data Services and Solutions (ScaDS), Ritterstrasse 9-13, 04109 Leipzig, alkhouri@informatik.uni-leipzig.de

² Competence Center for Scalable Data Services and Solutions (ScaDS), wilke@informatik.uni-leipzig.de

³ Der Quellcode der beschriebenen Analysen und weitere Arbeiten finden sich unter <https://github.com/GeorgesAlkhouri/golddust>.

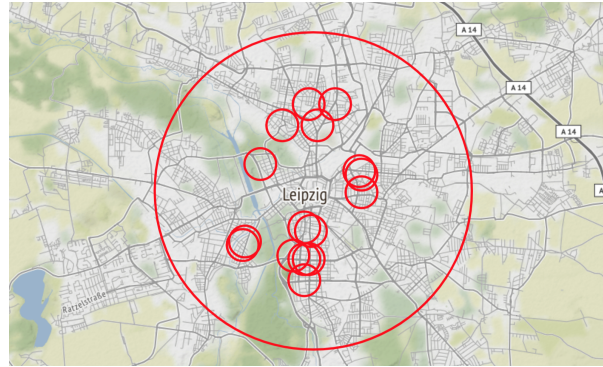


Abb. 1: Sensoren in Leipzig

sind, wurden die validen Messwerte bereinigt und die mittleren Werte für Temperatur, Luftfeuchtigkeit und Feinstaubbelastung von allen Sensoren in Leipzig zu einer Zeitreihe zusammengefasst. Hierbei wurden stündliche Intervalle aggregiert.

3 Zeitreihenanalyse und Vorhersage

Die extrahierte Zeitreihe wurde verwendet um ein neuronales Netz zu trainieren. Zur Vorhersage der Zielvariable PM10 wurden verschiedene Merkmale aus dem Datum und der angegebenen Zeit des Sensors genutzt und zudem weitere Merkmale gebildet. Prototypisch wurden die Merkmale Stunde, Tag des Monats und Monat genutzt und die Merkmale Wochentag, Wochennummer und Jahreszeit konstruiert. Des Weiteren fließen die durchschnittliche Temperatur und Luftfeuchtigkeit, in einer stündlichen Auflösung, in die Vorhersage ein.

Der verwendete neuronale Netz-Prototypen, besteht aus einer LSTM-Schicht [HS97] mit 48 Neuronen. Um einem Overfitting vorzubeugen wurde eine regulierende Dropoutschicht hinzugefügt und anschließend die Vorhersage von einem linearen Neuron durchgeführt. Hierbei konnte mit diesem simplen neuronalen Netz und den weiter oben aufgeführten Merkmalen eine mittlere quadratische Abweichung von $32,232 \mu\text{g}/\text{m}^3$ erreicht und ein erster Eindruck der Vorhersageleistung gewonnen werden, welcher in Abb. 2 dargestellt ist.

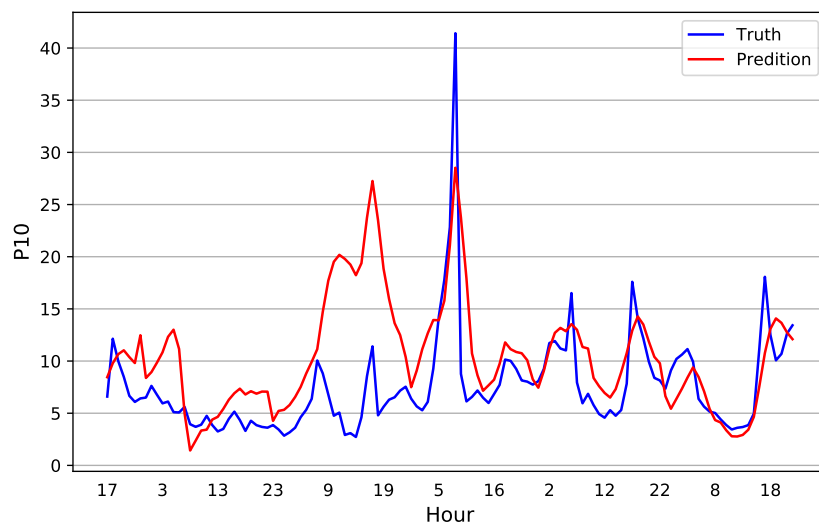


Abb. 2: Erste Validierung der Vorhersage mit einem MSE von $32.232 \mu\text{g}/\text{m}^3$ durch ein LSTM Netz.

Obwohl der illustrierte Zeitabschnitt, in Abb. 2, dem neuronalen Netz zum Trainingszeitpunkt nicht bekannt war, ist ein Vorhersagbarkeit tendenziell zu erkennen.

4 Weitere Arbeiten

Um die Aussagekraft der extrahierten Zeitreihe zu erhöhen, soll diese in ein Walk-Forward-Modell transformiert werden. Dies bedeutet, dass der Erwartungswert PM10 nach jeder Vorhersage als Merkmal für die nächste Vorhersage zur Verfügung steht. Formal soll dies folgendes Beispiel illustrieren:

$$\begin{aligned} (x_0, \dots, x_n)_t &\rightarrow PM10_t \\ (x_0, \dots, x_n, PM10_t)_{t+1} &\rightarrow P; 10_{t+1} \end{aligned}$$

, wobei $x_0, \dots, x_n$ Merkmale zur Vorhersage von PM10 darstellen und t den Zeitschritt definiert. Weiter gilt es, die Zeitreihe unter dem Erwartungswert PM2,5 zu untersuchen und auszuwerten.

Da exemplarisch nur ein Standort berechnet wurde, bietet sich die Möglichkeit, die Vorhersage auf mehrere Sensoren auszudehnen. Hierbei könnten dann weitere Faktoren wie Verkehrsdaten, Bevölkerungsdichte und Kategorisierung der Umgebung (Wohn-, Industrie- oder Naherholungsgebiet) in die Vorhersage einfließen. Hierbei bietet sich auch an zu prüfen, ob und wie entsprechende Daten aus öffentlichen Quellen generiert werden können. So gibt es in vielen Städten Webcams zur Betrachtung der aktuellen Verkehrslage die auswerten ließen. Abschließend ist eine Validierung der Qualität der PM10 Messwerte und Vorhersagen mittels der Daten des Bundesumweltamtes⁴ notwendig.

⁴ <https://www.umweltbundesamt.de/daten/luft/feinstaub-belastung>

5 Verwendete Technologien

Zur Verarbeitung der Rohdaten wurden IBM Cloud SQL Queries in Kombination mit Jupyter Notebooks verwendet. Die Vorhersage basiert auf der Deep-Learning Bibliothek Keras [Ch15]. Es wurde prototypisch nur ein (kombinierter) Sensor untersucht. Für die parallele Echtzeit-Verarbeitung von mehreren Sensoren bietet sich ein cloud-basiertes Streaming Framework, bspw. Apache Flink⁵ oder Spark⁶ an.

Literaturverzeichnis

- [Bu18] Budde, Matthias; Müller, Thomas; Laquai, Bernd; Streibl, Norbert; Schwarz, Almuth; Schindler, Gregor; Riedel, Till; Beigl, Michael; Dittler, Achim: Suitability of the Low-Cost SDS011 Particle Sensor for Urban PM-Monitoring. In: 3rd International Conference on Atmospheric Dust. 2018.
- [Ch15] Chollet, François et al.: Keras. <https://keras.io>, 2015.
- [HS97] Hochreiter, Sepp; Schmidhuber, Jürgen: Long Short-term Memory. Neural computation, 9, 1997.
- [Or06] Organization, World Health: Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization, 2006.

⁵ <https://flink.apache.org/>

⁶ <https://spark.apache.org/>