

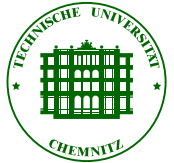
Abschätzung der Vertrauenswürdigkeit von Neuronalen Netzprognosen bei der Prozessoptimierung



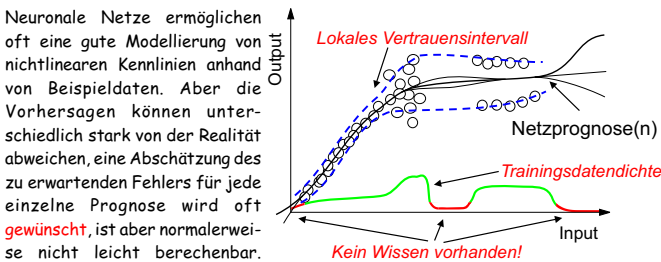
Peter Protzel Lars Kindermann Michael Tagscherer Achim Lewandowski

FORWISS
Bayerisches Forschungszentrum für wissenschaftliche Systeme
Forschungsgruppe Neuronale Netze und Fuzzy Logik, Erlangen
www.forwiss.de/~aknn
kindermann, tagscherer, lewandowski@forwiss.de

TU Chemnitz
Fakultät für Elektro- und Informationstechnik
Institut für Automatisierung
www.infotech.tu-chemnitz.de/~proauf
peter.protzel@e-technik.tu-chemnitz.de



Neuronale Netze als universale Funktionsapproximatoren liefern zu jeder beliebigen Kombination von Eingangsgrößen stets ein Resultat, unabhängig davon, wie erfolgreich das Training war. Auch wenn die aktuellen Eingangsgrößen in den vorhandenen Trainingsdaten überhaupt keine Repräsentation haben, wird das Netz ein beliebiges, bestenfalls zweifelhaftes Ergebnis anzeigen. Mit zusätzlichen Netzen, die neben der Prognose eine parallele Ausgabe von zu erwartendem Fehler und der Vertrauenswürdigkeit liefern, können bei Optimierungsaufgaben mögliche Fehleinstellungen erkannt und vermieden werden.



Setzt man ein Neuronales Netz als Streckenmodell in einer Regelung oder zur Optimierung von Prozessparametern ein, sollte man vermeiden, Arbeitspunkte anzu-steuern, für die das Netz keine vertrauenswürdigen Prognose liefern kann.

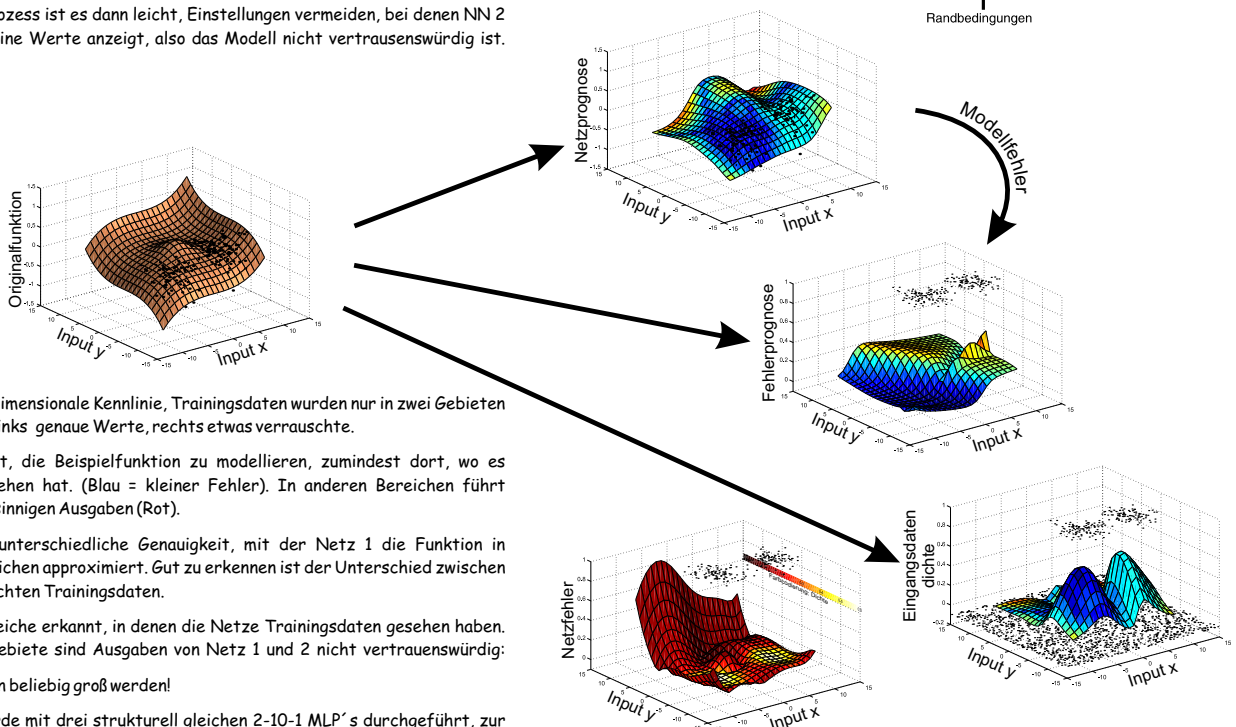
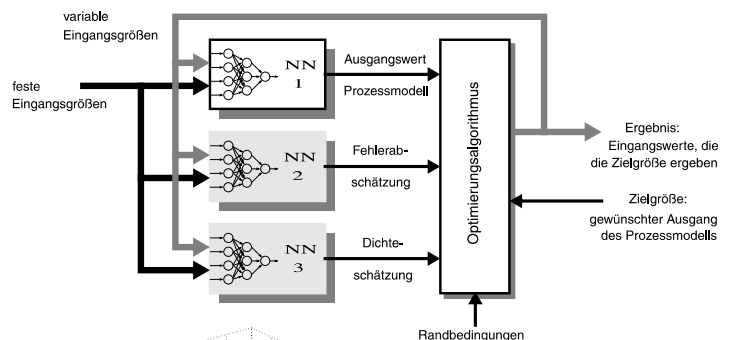
Eine einfacher Lösungsansatz

NN 1 ist ein Neuronales Netz, das mit bekannten Eingangs- und dazugehörigen Ausgangsgrößen aus dem zu modellierenden Prozess trainiert wird.

NN 2 hat die gleiche Struktur und wird mit denselben Eingangswerten wie NN 1 trainiert, aber mit dem Absolutfehler von NN 1 als Zielgröße. Es liefert damit ein lokales Maß für die Modellgenauigkeit im Trainingsbereich.

NN 3 wird ebenfalls mit denselben Eingangsgrößen wie NN 1 trainiert aber mit konstant "1" als Zielgröße. Außerdem zusätzlich mit zufällig erzeugten Inputs, die den gesamten möglichen Eingangsraum abdecken und "0" als Zielgröße. Dieses Netz liefert große Werte am Ausgang, wenn in der Umgebung der aktuellen Eingangsgröße viele Trainingsdaten lagen und kleine Werte, wenn kein entsprechendes Vorwissen vorhanden ist, also eine Approximation der Dichte im Eingangsraum.

Für den Optimierprozess ist es dann leicht, Einstellungen vermeiden, bei denen NN 2 große und NN 3 kleine Werte anzeigt, also das Modell nicht vertrauenswürdig ist.



Beispiel für eine 2-dimensionale Kennlinie, Trainingsdaten wurden nur in zwei Gebieten präsentiert: Vorne links genaue Werte, rechts etwas verrauschte.

Netz 1 hat gelernt, die Beispielfunktion zu modellieren, zumindest dort, wo es Trainingsdaten gesehen hat. (Blau = kleiner Fehler). In anderen Bereichen führt Extrapolation zu unsinnigen Ausgaben (Rot).

Netz 2 zeigt die unterschiedliche Genauigkeit, mit der Netz 1 die Funktion in verschiedenen Bereichen approximiert. Gut zu erkennen ist der Unterschied zwischen "guten" und verrauschten Trainingsdaten.

Netz 3 hat die Bereiche erkannt, in denen die Netze Trainingsdaten gesehen haben. Außerhalb dieser Gebiete sind Ausgaben von Netz 1 und 2 nicht vertrauenswürdig:

Der Netzfehler kann beliebig groß werden!

Die Berechnung wurde mit drei strukturell gleichen 2-10-1 MLP's durchgeführt, zur Dichteschätzung wurden 10 mal soviele "Nullvektoren" wie Trainingsdaten verwendet.