

Allgemeine Technische Bedingungen für NeuroModel Studien

Modelle und Modellgültigkeit

Der Auftraggeber und die atlan-tec Systems GmbH (im nachfolgenden ats genannt) vereinbaren zu Beginn der Modellierungsarbeiten, welche Daten aus welchen Zeiträumen für die Modellierung und für die Abnahme und Prüfung / Validierung der Modelle verwendet werden sollen.

Die vorliegenden Daten aus diesen Zeiträumen müssen vollständig sein und alle Online-Variablen aus dem PLS, sowie alle Labordaten enthalten, welche problemrelevant sind. Zudem informiert der Auftraggeber ats auf Anfrage verbindlich über alle betrieblichen Besonderheiten während dieser Zeiträume, wie insbesondere Abstellungen, Rohstoffänderungen, Produktwechsel und andere Änderungen im Prozess, welche nicht in den Datensätzen eindeutig erkennbar sind.

Auftraggeber und ats sind sich einig, dass in den vorliegenden Daten für die Modellierung nicht alle möglichen Betriebszustände des Prozesses enthalten sind. Als Ergebnis einer Modellierung werden daher auch nicht unbedingt alle möglichen Betriebszustände eines Prozesses in den Modellen abgebildet. Es werden nur für die Wertebereiche und damit Betriebszustände Modelle erstellt, welche in den erhaltenen Datensätzen in hinreichender Zahl und Güte zur Verfügung stehen. Welche Wertebereiche das sind, orientiert sich an dessen Untersuchungen zur Häufigkeitsverteilung der einzelnen Größen in den zu Grunde liegenden Datensätzen. In der Regel werden alle Datensätze entfernt, bei denen mindestens eine Variable in einem Bereich liegt, welcher in weniger als 90% der üblichen Betriebszustände gegeben ist. Dabei kommt es naturgemäß zur Einschränkung der Aussagefähigkeit und der Gültigkeit der Modelle.

Da die Charakteristik des Herstellungsprozesses verschiedener Produkte in einzelnen Produktionsanlagen unterschiedlich ist, ist es möglich, dass als Ergebnis dieser Studie die Separierung der Modelle nach Produkten empfohlen wird; d.h., dass für jedes Produkt ein eigenes Modell erstellt werden muss. Diese Separation in Teilmodelle verbessert die Genauigkeit der neuronalen Modelle erheblich und muss deshalb zur späteren praktischen Einsetzbarkeit des Systems gewährleistet sein bzw. durchgeführt werden. Die Erkennung des Betriebszustandes der Anlage oder eines charakteristischen Phasenbildes bei einem Batch und damit des hergestellten Produktes wird über die Identifikation unterschiedlicher Wertebereiche realisiert.

Als Ergebnis dieser Untersuchung werden daraus folgend nicht unbedingt für alle Produkte die notwendigen Modelle erstellt, sondern nur für repräsentative Produkte oder Wertebereiche. Es werden in der Regel im Rahmen einer Machbarkeitsstudie wie der hier angebotenen Arbeit nur für die Wertebereiche und damit Modelle für Produkte erstellt, welche in den erhaltenen Datensätzen in hinreichender Zahl und Güte zur Verfügung stehen. Welche Wertebereiche das sind, obliegt dem jeweiligen Projektierer, orientiert sich aber primär an dessen Untersuchungen zur Häufigkeitsverteilung der einzelnen Größen in den zu Grunde liegenden Datensätzen. In der Regel werden die Produkte untersucht, die jeweils mindestens 20% der Produktionszeit ausmachen. Dabei kommt es zur Eingrenzung der Aussagefähigkeit und der Gültigkeit der Modelle, was in der Regel dazu führen wird, dass nur gleiche oder ähnliche Betriebszustände der Anlage mit dem Modell betrachtet werden können, welche Grundlage des Trainings waren.

Die Erweiterung der Modelle erfolgt während der ggf. später erfolgenden Online-Implementierung eines Modells, bei der in der Regel dann das Online-Trainingsmodul *NeuroModel® Script Online* eingesetzt wird, um die Modelle regelmäßig mit neuen Daten und Betriebsbereichen zu aktualisieren.

Da die Neuronale Modellierung wie jede statistische Methode zur Modellbildung kein deterministisches Verfahren ist, welches physikalische Gegebenheiten berücksichtigt, sondern sich innerhalb des Modells immer auf einheitenlose und physikalisch aussageleere Größen bezieht, welche lediglich reproduzierbar sein müssen, hat der Austausch eines Sensors oder eines Analysators, die Änderung eines Wertebereichs oder aber die Veränderung eines Verfahrensschrittes gravierende Auswirkungen auf die Richtigkeit und Validierbarkeit des Modells. In der Regel muss nach einer Änderung der Abläufe, Variablen oder Sensoren, welche dem Modell zu Grunde liegen, das Modell durch die Einführung von Skalierungsfaktoren angepasst oder neu erstellt werden. Wird daher die Studie oder ein Projektierungsschritt mit Datenbeständen durchgeführt, welche auf einer Konfiguration der Anlage- und der Sensoren basiert und es werden dann nach Abschluss der Arbeit (Modellerstellung) Änderungen dieser Parameter durchgeführt, kann das Modell unter bestimmten Umständen ungültig werden. Dies führt zu einer Notwendigkeit der kostenpflichtigen Wiederholung eines Teils der durchgeführten Arbeiten führt.

Kommt es zu einer Änderung der Wertebereiche einzelner Größen, deren Messverfahren oder zu einer Verfahrensänderung, müssen diese Änderungen zum Zeitpunkt der Durchführung ats mitgeteilt werden, damit keine Arbeiten durchgeführt werden, die zu unbrauchbaren Ergebnissen führen. Unterlässt der Auftraggeber diese Anzeige an ats kommt es zur Mehrung. Ändert sich das Verfahren in dieser Weise nach Abschluss einer Projektphase, ist die Anpassung oder Wiederholung derselben jeweils mit Mehrkosten für den Auftraggeber verbunden.

Aussagen, welche aus den erstellten Modellen abgeleitet werden, sind nur in der Nähe tatsächlicher Betriebsbereiche richtig. Je weiter das Modell sich von tatsächlich bei der Modellierung berücksichtigten Betriebsbereichen entfernt, desto unsicherer werden die Aussagen des Modells bezüglich Sensitivität und Kausalität.

Ist die Anzahl und die Dichte der gültigen und in der Modellierung verwendeten Datensätze geringer, als statistisch erforderlich, ist eine hohe Qualität der Modelle nicht sicher gewährleistet. Diese Tatsache ist nicht auf einen Mangel der Modellierungsmethode zurück zu führen, sondern auf einen Mangel der zur Verfügung gestellten Datensätze.

Modelle, welche von ats erfolgreich erstellt und durch den Auftraggeber abgenommen wurden, verlieren im Moment der Änderung des Verfahrens, der Änderung der Wertebereiche bestimmter Variablen, der Änderung von Messprinzipien oder von Position von Messwertaufnehmern, sowie im Moment der Unterlassung der Anzeige einer solchen Änderung durch den Auftraggeber die Gewährleistung.

Abnahmen beziehen sich immer nur auf Zeiträume vor einer Änderung des Verfahrens, der Wertebereiche bestimmter Variablen, der Messprinzipien oder der Position von Messwertaufnehmern. Ändert sich nach der Ausführung der Arbeiten durch ats ein Parameter derart, dass eine Abnahme mit den neuen Datensätzen unmöglich ist, werden diese Datensätze nicht zur Abnahme verwendet.

In diesem Fall sind nur Datensätze aus Zeiträumen vor der Verfahrensänderung relevant.

Benötigte Daten aus dem Prozess

Für die Modellierung des Prozesses mit Neuronalen Prozessmodellen werden gleichzeitige und mit Zeitstempel aufgezeichnete äquidistante Datensätze aus dem Prozess, sowie Laboranalysen benötigt. Ersatzweise kann der Auftraggeber Datensätze mit fertigen Fingerprints von Batches bereitstellen.

Sofern Laboranalysen gezogen werden, welche Zwischengrößen oder Zielgrößen der Modellierung umfassen, ist es notwendig, den Zeitpunkt der Probenahme möglichst genau feststellen zu können. Sofern hier Ungenauigkeiten angenommen werden, sollten diese möglichst genau quantifiziert werden.

Es wird hierzu empfohlen, diese Probenahme nicht zu festgelegten Zeitpunkten anzuordnen, sondern innerhalb eines bestimmten Zeitintervalls, jedoch mit der Anweisung, dass der Zeitpunkt der Probenahme genau notiert werden muss. Durch diese Maßnahme kann in der Regel eine höhere Genauigkeit des notierten Zeitpunktes festgestellt werden, da Fehlnotierungen weniger häufig sind.

Datenformate

Um eine Studie mit Neuronalen Netzen durchzuführen, werden Prozessdaten und Labor- oder Analysatordaten benötigt. Dieses stellt der Auftraggeber der ats zur Verfügung. Für vom Auftraggeber gestellten und an ats übergebene Daten gelten dabei folgende Grundanforderungen.

Abweichungen der Datensätze von diesen Vorgaben sind mehrerungsrelevant, verursachen also Mehrkosten, die nach Sichtung der Daten von ats an den Auftraggeber berechnet werden, da die Umformatierung mit teils erheblichem Aufwand verbunden ist. Diese Grundanforderungen gelten völlig unabhängig vom übergebenen Datenformat:

- Die Spalten (Felder) beinhalten je Spalte oder Feld genau eine Variable oder Zeit/ Datum
- Die Zeilen (Datensätze) beinhalten je Zeile oder Datensatz einen Wert für jede Variable oder für Zeit und Datum, die den richtigen Spalten/ Feldern mit den zugehörigen Variablennamen immer gleichartig und richtig zugeordnet sind. (Datenlücken sind zulässig, der Wert ist dann NULL, nicht aber Doppelbesetzungen).
- Den Datenzeilen muss eine Zeile mit Spaltennamen vorangehen
- Den Datenzeilen darf eine Zeile mit den physikalischen Einheiten der Spalten vorangehen.
- Spaltennamen dürfen nicht doppelt vergeben werden. Jeder Spaltenname und damit Variablenname darf in der gesamten Datenbasis nur einmal vorkommen.
- Das Datum hat folgendes (in Deutschland übliche) Format: "dd.mm.yyyy hh:mm:ss" (Sekunden und führende 0 bei Stunde, Tag und Monat sind optional; die Jahreszahl muss vierstellig sein und die Stunde ist auf der 24-Stunden-Skala anzugeben).
- Zahlenwerte, die leere Zellen markieren sollen (wie z. B. -99.999 oder gar 0) sind **nicht** zugelassen. Leere Zellen bleiben einfach leer. Beträgt der Messwert tatsächlich 0, wird natürlich eine 0 eingetragen.
- Alle Daten, also alle Prozessdaten und alle Labordaten, müssen (unabhängig vom Format) in einer Datei bzw. dort in einer Tabelle vorliegen.

- Die Labordaten sind jeweils in einer separaten Spalte, aber in derselben Tabelle in einer Zeile einzugeben, in der auch alle anderen Messdaten vorhanden sind. Liegen die Labordaten nicht exakt in demselben Zeitraster wie die Messdaten vor, sind diese in die Messwertzeile einzutragen, die der Probenahme zeitlich am nächsten ist
- Die Zeitabstände zwischen den Zeilen müssen äquidistant sein, das bedeutet, die Erfassung der Messwerte muss in gleichmäßigen Intervallen erfolgen und auch so in den Datensätzen abgelegt werden

Aufwand, der aufgrund fehlerhafter Datenträger oder Dateien entsteht, kann zu Mehrung führen.

Sofern **Laboranalysen** gezogen werden, welche Zwischengrößen oder Zielgrößen der Modellierung umfassen, ist es notwendig, den Zeitpunkt der Probenahme genau feststellen zu können. Es wird empfohlen, diese Probenahme nicht zu festgelegten Zeitpunkten anzuordnen, sondern innerhalb eines bestimmten Zeitintervalls, jedoch mit der Anweisung, dass der Zeitpunkt der Probenahme genau notiert werden muss. Durch diese Maßnahme kann in der Regel eine höhere Genauigkeit des notierten Zeitpunktes festgestellt werden, da Fehlnotierungen weniger häufig sind.

Die Verwendung von **Datenerfassungssystemen** mit so genannten „Kompressionsverfahren“, wie die „PIMS“-Systeme von Aspen (Infoplus 21) oder OSI (PI-System) führt dazu, dass die Daten unter Umständen für jede systematische Verwendung unbrauchbar werden. Dadurch kann eine Modellierung unmöglich werden. Werden Daten aus diesen Systemen verwendet, müssen deren „Kompressionsverfahren“ **vollständig** abgeschaltet werden.

Für übergebene Daten gelten folgende Anforderungen:

Alle Dateien und Tabellen werden in einem beigefügten Dokument (Textfile oder gleiches Format wie die Daten) genau beschrieben, d. h. Inhalt und Struktur sind detailliert dokumentiert.

Die übergebenen Datensätze müssen – soweit dies nicht gemäß Leistungsverzeichnis in der Verantwortung von ats liegt – eines der folgenden Formate haben:

Access-Format (nur zulässig sind: Access2007 bis Access2010)

- Der Spaltentyp darf nur "Date" oder "Double" sein. In Ausnahmefällen ist für reine Texte der Typ "Text" zugelassen!

Excel-Formate (nur zulässig sind: Excel 2007 bis Excel 2010)

- Die Datenzellen müssen alle als "Zahl" (Wert) definiert werden
- Die Datei darf nur Daten enthalten und keine Grafiken, Bitmaps oder ähnliche Sonderstrukturen, welche mit den Daten nichts zu tun haben
- Alle bestehenden Formeln müssen durch numerische Werte ersetzt werden.
- Die Dateien müssen „frisch“ geschrieben werden, da mehrfaches Löschen und erweitern in Excel-Tabellen diese instabil macht. Die beste Methode dazu ist: Kopieren der Datenfelder in die Zwischenablage und „Inhalte Einfügen“ als „Wert“ in eine völlig neue Excel-Tabelle ohne weitere Inhalte und speichern dieser neuen Excel-Tabelle unter neuem Dateinamen.
- Besondere Vermerke zu jeder Tabelle dürfen in der Zelle A1 durch die "Kommentar"-Funktion eingetragen werden.
- Besondere Vermerke zu den Spalten dürfen in den jeweiligen Spaltenköpfen durch die "Kommentar"-Funktion in die Tabelle eingetragen werden.
- Die Verwendung der „Kommentar“-Funktion an anderen Stellen ist nicht zulässig.
- Ebenfalls unzulässig ist es, wichtige Besonderheiten und Auffälligkeiten (allein) durch Zellformatierung zu definieren. Soll beispielsweise auf einen Zeitraum mit fehlerhaften Daten hingewiesen werden, sind diese Daten zu löschen und keinesfalls durch rote Einfärbung oder dergleichen zu markieren. Die Zellformatierung geht bei der Bearbeitung der Daten verloren und darf deshalb keine wesentliche Information enthalten.

ASCII-Format

- Erforderliches Spaltentrennzeichen: Semikolon. Alle Zeilen müssen grundsätzlich die gleiche Anzahl an Semikolons aufweisen. Punkte oder Kommata als Trennzeichen dürfen niemals verwendet werden, da sonst eine Verwechslung mit Tausenderzeichen oder Dezimaltrennzeichen auftreten kann.
- Ersatzweise kann als Spaltentrennzeichen SPACE oder TABULATOR verwendet werden. Die Verwendung beider Trennzeichen zusammen oder im Wechsel ist unzulässig und führt zu Unbrauchbarkeit der Dateien oder zu Mehrung. Alle Zeilen müssen bei TABs grundsätzlich immer die gleiche Zahl TABs enthalten; zwischen zwei Spalten darf nur ein TAB enthalten sein. Bei SPACES können zwischen zwei Spalten auch mehrere SPACES enthalten sein – aber auch nur dann, wenn außer den SPACES kein anderes Trennzeichen vorhanden ist und keine leeren Zellen vorkommen.
- Jede Zeile ist ein Datensatz und muss die gleiche Anzahl Spalten (Zahlen) aufweisen.
- Leere Zellen (also fehlende Werte) werden beim Schreiben ausgelassen, so dass bei Einhaltung der Spezifikation („;“ als Trennzeichen) ein “;” lesbar ist. Irgendwelche Zahlenwerte, die leere Zellen markieren sollen (wie z.B. -99.999) sind bei Semikolon als Trennzeichen nicht zugelassen.
- Als Dezimaltrennzeichen dient in der Regel das Komma.
- Tausendertrennzeichen dürfen nicht verwendet werden
- Jede Abweichung von Verwendung des Kommas als Dezimaltrennzeichen oder des Semikolons als Feldseparator ist deutlich zu dokumentieren.

Andere Formate

Alle weiteren Formate wie z.B. FoxPro sind nicht immer importierbar, so dass z.T. zusätzlicher erheblicher Aufwand auftreten kann. Datenbanken, sofern sie importierbar sind, unterliegen grundsätzlich denselben Mindestregeln wie Access-Datenbanken.

Erfüllen die übergebenen Datensätze nicht die oben beschriebenen Anforderungen, ist in der Regel mit erheblichem Mehraufwand zu rechnen.