

DAV³E: Data Analysis and Verification/Visualization/Validation Environment für die Multisensor-Datenfusion

Manuel Bastuck, Tobias Baur, Andreas Schütze
Lehrstuhl für Messtechnik, Campus A5.1, Universität des Saarlandes, 66123 Saarbrücken

Zusammenfassung

DAV³E, eine vollständig objektorientierte MATLAB-Toolbox zur Auswertung zyklischer Sensordaten, wird vorgestellt. Es können beliebig viele Daten importiert und individuelle Vorschriften zur Vorverarbeitung und Merkmalsextraktion definiert werden. Aus den Merkmalen kann anschließend mit vordefinierten Dimensionsreduktions- und Klassifizierungsalgorithmen ein Modell erzeugt werden, das in der Lage ist, neue, unbekannte Daten zu klassifizieren. Der graphische Aufsatz für die Toolbox leitet den Anwender durch die Auswertung und stellt umfangreiche Funktionen zur Visualisierung zur Verfügung, so dass neue Daten einfach untersucht werden können, um ein optimales Ergebnis zu erreichen. Im Falle einer bereits vorhandenen Auswertestrategie kann die kommandozeilenbasierte Version der Toolbox genutzt werden, um die Auswertung zu automatisieren, oder verschiedene, neue Methoden schnell zu validieren.

Keywords: Software, Mustererkennung, Datenauswertung, Visualisierung, Automatisierung

Einführung

Im Gegensatz zum statischen Betrieb von Sensoren kann deren Empfindlichkeit, Selektivität und Stabilität oft durch Betrieb mit zyklisch variierten Parametern verbessert werden. Beispiele dafür sind der temperaturzyklische Betrieb (temperature cycled operation, TCO) [1,2] bzw. die Impedanzspektroskopie bei Gassensoren [3], aber auch Sensorsysteme zur Überwachung von Hydraulikanlagen [4], die den zyklischen Arbeitsrhythmus dieser Anlagen ausnutzen. Jeder Zyklus stellt dabei einen Datenpunkt im n-dimensionalen Zustandsraum dar, wobei n die Anzahl der Datenpunkte pro Zyklus ist. Die Dimensionalität kann und muss [5] drastisch reduziert werden, indem aus dem Zyklus signifikante Merkmale errechnet werden. Merkmale reichen von einfachen mathematischen Vorschriften, z.B. dem Mittelwert über ein bestimmtes Intervall des Zyklus, über Zeitkonstanten und statistische Momente bis hin zur Bestimmung modellbasierter Parameter aus einer hierfür optimierten Zyklusform [6]. Zusätzlich erhöhen Varianten der Vorverarbeitung der Rohdaten und Merkmale, des Algorithmus zur weiteren Dimensionsreduzierung (Hauptkomponentenanalyse, PCA [7], Lineare Diskriminanzanalyse, LDA [8], u.a.) sowie des Klassifikators (k-nächste-Nachbarn [8], Mahalanobisdistanz [9],

neuronale Netze, Support Vector Machines, SVM [10], u.a.) oder des Regressionsalgorithmus (Partial Least Squares Regression, PLSR [11], Support Vector Regression, SVR [12], u.a.) die Komplexität drastisch, so dass die optimale Auswertung solcher Daten schnell sehr komplex und zeitaufwändig wird. In Abb. 1 ist der Weg von den Rohdaten zum Modell nochmals dargestellt.

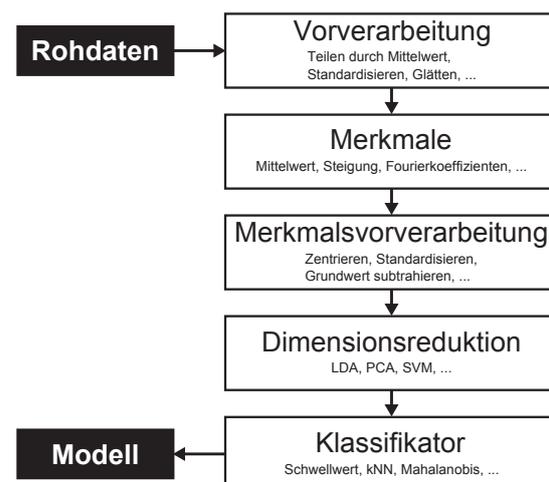


Abb. 1: Der Weg von zyklischen Rohdaten zu einem Modell, das neue Daten korrekt klassifizieren kann. Jeder Zwischenschritt wirkt als Multiplikator für die Anzahl möglicher Modelle.

Jeder Schritt fungiert als Multiplikator für die Anzahl der möglichen Modelle. In den ersten drei Schritten – Merkmale und (Merkmals)Vorverarbeitung – können alle jeweils verfügbaren Optionen beliebig miteinander kombiniert werden, so dass mit jeder neuen Option, bspw. einem neuen Verfahren zur Vorverarbeitung, die Anzahl der Möglichkeiten allein für diesen einen Schritt sehr schnell ansteigt. Werden in einem System verschiedene Sensoren verwendet, so kann der in Abb. 1 dargestellte Ablauf für jeden dieser Sensoren unterschiedlich sein. Da in vielen Fällen keine Theorie existiert, auf deren Basis ein Auswertungsweg einem anderen vorgezogen werden könnte, ist es wichtig, „schlechte“ Wege möglichst früh zu erkennen, und die verbleibenden Wege effizient validieren zu können.

Unsere vollständig objektorientierte MATLAB-Toolbox DAV³E (Data Analysis and Verification/Visualization/Validation Environment) ermöglicht den einfachen und effizienten Umgang auch mit sehr großen Datenmengen, die aus beliebig vielen Dateien zusammengestellt werden können. Bei der Entwicklung wurde großer Wert auf Modularität und einfache Erweiterbarkeit gelegt, so dass neue Funktionen für Vorverarbeitung, Merkmalsberechnung usw. einfach implementiert werden können. Die grafische Benutzeroberfläche (graphical user interface, GUI) erlaubt einfache Bedienung und direkte Visualisierung der Daten, die kommandozeilenbasierte Verwendung der Toolbox kann hingegen für die automatisierte Validierung vieler verschiedener Abläufe oder zur Batch-Verarbeitung vieler einzelner Messdateien genutzt werden.

Interne Datenstruktur

Beim Import von Daten in DAV³E werden diese in eine Baumstruktur aus „Projekt“, „Messungen“, „Clustern“ und „Sensoren“ zerlegt. Der Import erfolgt vorzugsweise aus Dateien im open-source-Format HDF5 (Hierarchical Data Format, [13]), welches zur strukturierten und effizienten Speicherung sehr großer Datenmengen entwickelt wurde. Es setzt ebenfalls auf eine Baumstruktur, in der jedes Element mit Metadaten („Attributen“) versehen werden und blockweise ausgelesen werden kann, wodurch auch mehrere Gigabyte im Optimalfall in wenigen Sekunden eingelesen werden können. Grundsätzlich können jedoch Daten aus jeder dokumentierten Quelle eingelesen werden.

Ein „Cluster“ enthält logische Gruppierungen von „Sensoren“ bzw. Sensorsignalen, beispielsweise die angelegte Spannung, gemessener Strom und interne Zeit eines

Messsystems. Sensorsignale in demselben Cluster wurden per Definition mit derselben Samplerate und ohne Offset untereinander aufgenommen, was eine besonders einfache Handhabung der Daten innerhalb desselben Clusters ermöglicht.

Mehrere Cluster mit potentiell unterschiedlichen Sampleraten und Offsets werden in einer „Messung“ zusammengefasst. Diese definiert für jeden Zeitpunkt die Umgebungsbedingungen bzw. Zielgrößen des Messablaufs, also bspw. die Konzentration verschiedener Gase. Mit diesen Informationen wird später für das Training der Algorithmus zur Dimensionsreduktion und des Klassifikators bestimmt, welche Sensorzyklen in welchen Umgebungsbedingungen stattgefunden haben. Offsets und Sampleraten der einzelnen Cluster werden dabei beachtet.

Ein „Projekt“ umfasst beliebig viele Messungen, und damit die Gesamtheit aller Daten. Für die spätere Merkmalsberechnung werden alle Messungen als eine einzige, längere Messung aufgefasst, d.h. die Messungen werden „seriell“ fusioniert, was die Gesamtanzahl der Zyklen erhöht. Die in jeder Messung enthaltenen Sensoren werden hingegen „parallel“ fusioniert, d.h. jeder Sensor liefert zusätzliche Merkmale für „seinen“ Teil der Gesamtmessung.

Benutzeroberfläche und Module

Der generelle Aufbau der Benutzeroberfläche ist in Abb. 2 zu sehen. Links sind alle geladenen Module in der Reihenfolge gelistet, die für einen typischen Arbeitsablauf üblich und sinnvoll ist. In jedem der Module wird ein genau definierter Teilbereich des Gesamtprojekts bearbeitet, und ein Modul kann erst ausgewählt werden, wenn alle zuvor notwendigen Schritte bereits durchgeführt wurden. Bspw. kann das Modul zur Definition der Zielvektoren erst aufgerufen werden, wenn mindestens ein Bereich in der Messung markiert wurde, für den ein Eintrag im Zielvektor erstellt werden kann. Durch diese klare Struktur kann DAV³E bereits nach kurzer Einarbeitung selbstständig bedient werden.

Weiterhin stehen Optionen zum Laden und Speichern des Projekts oder Teilen davon, eine Historie der ausgeführten Aktionen, sowie die Auswahl des aktuell angezeigten Sensorsignals immer und unabhängig vom gerade aktiven Modul, zur Verfügung. Diese Funktionen sind Teil der Haupt-GUI, die einen Rahmen für die einzelnen Module zur Verfügung stellt. Dadurch können neue Module für spezifische Anwendungen leicht erstellt und eingefügt werden.

Verfügbare Module und Arbeitsablauf

Die derzeit in der Toolbox enthaltenen Module erlauben es, zyklische Daten beliebig vieler Sensoren darzustellen, Vorverarbeitungen anzuwenden und Merkmale zu berechnen, sowie diese Merkmale zu inspizieren und daraus mit verschiedenen Algorithmen Modelle zur Klassifizierung neuer Daten zu berechnen. Im Folgenden wird auf die einzelnen Module, ihre Funktion, Bedienung und Besonderheiten, näher eingegangen. Derzeit existieren neun Module: Import, Link sensors, Preprocessing, Select relevant cycles, Grouping, Define features, Compose, Visualization und Model.

Im Modul *Import* werden beliebig viele Messdateien geladen und einer Messung oder einem Cluster zugewiesen. Alle Sensoren werden in einer Tabelle angezeigt und können, zur besseren Übersichtlichkeit und für weniger Speicherplatzbedarf des Projekts, einzeln an- und abgewählt werden. Die Tabelle kann nach allen Spalten gefiltert werden, was insbesondere bei großen Projekten hilft, eine bestimmte Gruppe von Sensoren schnell zu finden. Zusätzlich können auch vordefinierte „virtuelle Sensoren“ aus einem oder mehreren realen Sensoren berechnet werden, z.B. der Sensor „Widerstand“ aus den tatsächlich gemessenen Sensoren „Strom“ und „Spannung“. Virtuelle Sensoren können anschließend genau wie reale Sensoren verwendet werden.

Im Modul *Link sensors* können Gruppen von Sensoren definiert werden. Dadurch müssen Einstellungen, z.B. eine bestimmte Vorverarbeitungssequenz, nicht für jeden dieser Sensoren einzeln vorgenommen werden, sondern sie können direkt auf die komplette Gruppe übertragen werden.

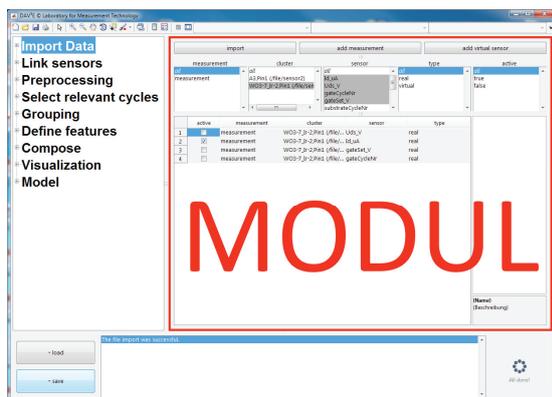


Abb. 2: Gesamtansicht der GUI. Der gekennzeichnete Bereich zeigt immer die Oberfläche des aktuellen Moduls. Die Bereiche außerhalb sind statisch und zeigen u.a. eine Liste aller geladenen Module (links).

Im Modul *Preprocessing* werden die geladenen Daten zum ersten Mal visualisiert. Hierfür stehen zwei sich ergänzende Ansichten zur Verfügung: zyklische und quasistatische Darstellung, zu sehen in Abb. 3 im unteren bzw. oberen Diagramm am Beispiel eines Gassensors. Unten werden die Zyklen zu den oben, d.h. im quasistatischen Signal, ausgewählten Zeitpunkten der Messung gezeigt. Es ist deutlich die Änderung in der Zyklusform zu sehen, die mit der Art des angelegten Gases korreliert. Gleichzeitig können in der zyklischen Ansicht (unteres Diagramm) feste Zeitpunkte in jedem Zyklus markiert werden, die – aufgetragen über die Zeit – die quasistatischen Signale im oberen Diagramm ergeben, in denen meist schon ein Einfluss des Gases bzw. anderer Umgebungsbedingungen festgestellt werden kann. Alle gewählten Punkte können zu jedem Zeitpunkt interaktiv verschoben werden; die Anzeige wird dabei sofort aktualisiert. In diesem Modul können auch Vorverarbeitungssequenzen aus verschiedenen, vordefinierten Funktionen (Standardisierung, Teilen durch Mittelwert, Glättung, etc.) zusammengestellt werden. Diese werden sofort auf die Daten angewendet und die so bearbeiteten Daten werden zusätzlich angezeigt. So ist auf einen Blick der Einfluss der gewählten Vorverarbeitung auf Zyklusform und quasistatisches Signal ersichtlich, was oft bereits eine erste Einschätzung über die Eignung der Vorverarbeitung erlaubt.

Im Modul *Select relevant cycles* (Abb. 4) werden die im vorigen Modul ausgewählten, quasistatischen Signale zur Orientierung angezeigt. In demselben Diagramm können dann „relevante“ Bereiche der Messung markiert werden. Die Relevanz hängt dabei von

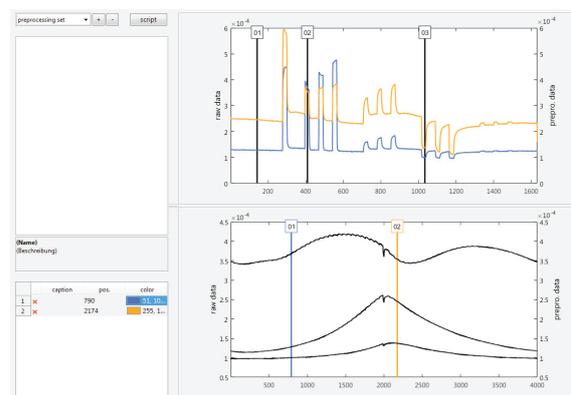


Abb. 3: Ansicht des Moduls „Preprocessing“. Oben ist das quasistatische Signal für die beiden unten gewählten Punkte (blau und orange) zu sehen, unten die Zyklen zu drei oben markierten Zeitpunkten in der Messung.

der Art der Daten ab. Im Falle eines Gassensors könnte es sich dabei zum Beispiel um Bereiche (also Zeitintervalle) der Messung handeln, in denen der Sensor verschiedenen Gasmischungen ausgesetzt war. Diese Bereiche können von Hand festgelegt oder aus einer vorhandenen Datei importiert werden. Außerdem können Metadaten in Form von „Komponenten“ hinzugefügt werden, um die Bereiche weiter zu charakterisieren, bspw. die jeweilige Konzentration einer Gaskomponente in ppm, oder die Schwere eines Fehlers in einer Hydraulikanlage in Prozent. Mithilfe vorgefertigter und erweiterbarer Skripte können die markierten Bereiche z.B. automatisch nach der Konzentration der Komponenten eingefärbt werden, verkleinert oder vergrößert werden, usw. Die Farbe der Bereiche wird in allen folgenden Schritten beibehalten, wenn z.B. Daten aus diesem Bereich visualisiert werden. Im Modul *Grouping* können für die zuvor festgelegten Bereiche der Messung beliebig viele Zielvektoren festgelegt werden. Jeder Vektor ordnet jedem Bereich eine eindeutige Bezeichnung zu. Diese Bezeichnungen geben die spätere Gruppierung der Bereiche für den Klassifizierungsalgorithmus vor. Zielvektoren können von Hand oder automatisch anhand der zuvor festgelegten Komponenten generiert werden. Zur Verdeutlichung sind in Tab. 1 drei beispielhafte Zielvektoren Vec1-3 dargestellt. Um Fehler bei der manuellen Festlegung der Zielvektoren schnell zu entdecken, werden die zuvor festgelegten Bereiche in dieser Ansicht noch einmal dargestellt, wobei Bereiche in der gleichen Gruppe in der gleichen Farbe dargestellt werden. Im Modul *Feature ranges* wird festgelegt, welche Merkmale zur Charakterisierung der

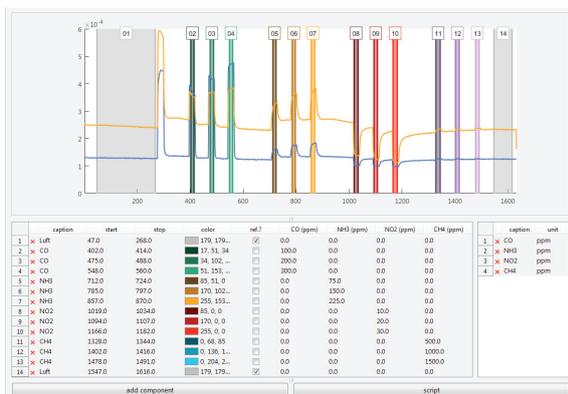


Abb. 5: Ansicht des Moduls „Select relevant ranges“. In den zuvor gewählten, quasistatischen Signalen können spezifische Bereiche der Messung markiert und farblich kodiert werden.

Tab. 1: Drei beispielhafte Zielvektoren. Vec1 definiert eine konzentrationsunabhängige Unterscheidung von CO und NH₃, während Vec2 und Vec3 die Konzentrationen nur jeweils eines Gases unterscheiden.

Bereich	Vec1	Vec2	Vec3
1	CO	100	ignore
2	CO	200	ignore
3	CO	300	ignore
4	NH3	ignore	75
5	NH3	ignore	150
6	NH3	ignore	225

Zyklen verwendet werden sollen. Zur Orientierung wird dabei, farblich kodiert, die mittlere Zyklusform in jedem der zuvor markierten Bereiche angezeigt (Abb. 5 unten), so dass besonders unterschiedliche Stellen im Zyklus leicht erkannt werden können. Der Benutzer hat die Auswahl aus vielen vordefinierten Berechnungsvorschriften (Mittelwert, Steigung, statistische Momente, Minimum, Maximum, etc.), und kann für jede dieser Funktionen beliebig viele Intervalle im Zyklus festlegen, auf denen das Merkmal, d.h. die jeweilige Funktion für jeden Zyklus berechnet werden soll. Das Merkmal wird für jedes Intervall und für jeden gemittelten Zyklus sofort berechnet und in einem zweiten Diagramm (Abb. 5 oben) dargestellt. Auf diese Weise ist schnell ersichtlich, ob sich das ausgewählte Merkmal zur Trennung eignet oder nicht. Für bestimmte Signalverläufe ist es oft sinnvoll, viele kleine Intervalle mit gleicher oder logarithmisch ansteigender Breite festzulegen. Um dies in der GUI zu vereinfachen, können große Intervalle nochmals automatisch in eine

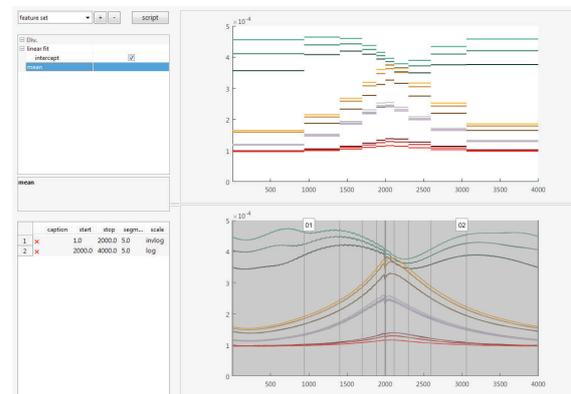


Abb. 4: Ansicht des Moduls „Feature ranges“, hier bei Auswahl des Mittelwerts als Merkmal. Für jede Funktion zur Merkmalsberechnung können beliebig viele Bereiche festgelegt werden, für die eine Vorschau des Merkmals

vorgegebene Anzahl von Abschnitten unterteilt werden, wie in Abb. 5 unten gezeigt. In diesem Modul können beliebig viele Merkmalssets, d.h. eine Zusammenstellung von Berechnungsvorschriften und dazugehörigen Intervallen, erzeugt werden, bspw. für unterschiedliche Sensoren, oder um unterschiedliche Ideen schnell testen zu können.

Im Modul *Compose* werden noch einmal alle Sensoren, die verfügbaren Vorverarbeitungssequenzen und Merkmalssets angezeigt. Aus dieser Auswahl kann nun ein multivariater Datensatz zusammengestellt werden, der anschließend für das Training des Modells benutzt wird. Jener kann alle oder nur eine Untermenge der insgesamt zur Verfügung stehenden Sensoren enthalten, wobei für jeden dieser Sensoren eine individuelle Vorverarbeitungssequenz und ein individuelles Merkmalsset definiert werden kann. Nach Fertigstellung der Auswahl werden alle ausgewählten Merkmale auf den entsprechend vorverarbeiteten Daten berechnet. Auch hier können beliebig viele dieser Datensätze erzeugt werden, so dass beim anschließenden Training schnell zwischen verschiedenen Datensätzen gewechselt werden kann.

Im Modul *Visualization* können die berechneten Merkmale auf verschiedene Weisen dargestellt werden. Zur Verfügung stehen unter anderem eine Darstellung der Merkmale über die Zeit (nützlich zur Erkennung von Drift), ein Boxplot (um die Trennkraft der Merkmale zu beurteilen) oder eine Korrelationsanalyse der Merkmale untereinander (um stark korrelierte Merkmale zu vermeiden). Auf die verschiedenen Darstellungsarten wird im Abschnitt „Visualisierung“ näher eingegangen.

Im Modul *Model* können die zuvor berechneten Datensätze aus Merkmalen schließlich genutzt werden, um eine Vorschrift zur Dimensionsreduktion zu finden, bspw. mittels PCA [7] oder LDA [8]. Das Ergebnis dieses Schritts kann, je nach Anzahl der verbleibenden Dimensionen, als Histogramm (pro Dimension) oder Streudiagramm (bis max. drei Dimensionen) dargestellt werden. In Kombination mit einem Klassifikator, bspw. k-nächste-Nachbarn (kNN) [8] oder Mahalanobis-Distanz [9], ergibt sich ein Modell zur Klassifizierung unbekannter Zyklen. Durch entsprechende Auswahl eines der zuvor definierten Zielvektoren kann das Modell für die Unterscheidung der dort festgelegten Gruppen optimiert werden. Das Modell wird im Anschluss an das Training z.B. durch k-fache Kreuzvalidierung überprüft, und das Ergebnis in Form einer Wahrheitsmatrix („confusion matrix“) und als Anteil der korrekt klassifizierten Daten in Prozent angezeigt.

Visualisierung

Grundlegende Muster, Fehler, Zusammenhänge und sonstige Strukturen sind besonders einfach bei geeigneter Visualisierung von Daten zu erkennen. Darum wird in DAV³E großer Wert auf die optische Darstellung aller Vorgänge gelegt.

Neben den bereits genannten Darstellungsformen für die Rohdaten wie zyklische und quasistatische Darstellung und Farbcodes zur Identifikation von einmal festgelegten Bereichen ermöglicht eine aussagekräftige Darstellung der Merkmale oftmals bereits eine gute Einschätzung der Trennkraft des endgültigen Modells.

Die Darstellung jedes Merkmals über die Zeit zeigt sofort, ob starke Sensordrift vorliegt – nämlich dann, wenn sich die Merkmalswerte eines Sensors über die Zeit – statt nur in Abhängigkeit der Umgebungsbedingungen – ändern. Gleichzeitig können in dieser Darstellung aber auch besonders interessante Merkmale identifiziert werden, bspw. solche, die sich nur in Anwesenheit einer bestimmten Komponente ändern.

Ein Boxplot fasst die Merkmale nach den im Zielvektor definierten Gruppen zusammen und stellt für jede dieser Gruppen den Mittelwert, die Standardabweichung und Ausreißer dar. Ein gutes Merkmal sollte eine geringe Standardabweichung innerhalb einer Gruppe aufweisen, und sein Mittelwert sollte sich für verschiedene Gruppen möglichst stark ändern. Eine andere Eigenschaft eines guten Merkmals ist eine starke Korrelation zur Zielgröße bei gleichzeitig geringer Korrelation mit allen anderen Merkmalen. Die Korrelation zur Zielgröße kann mit einem Wert pro Merkmal beschrieben und daher leicht in einem Linien-

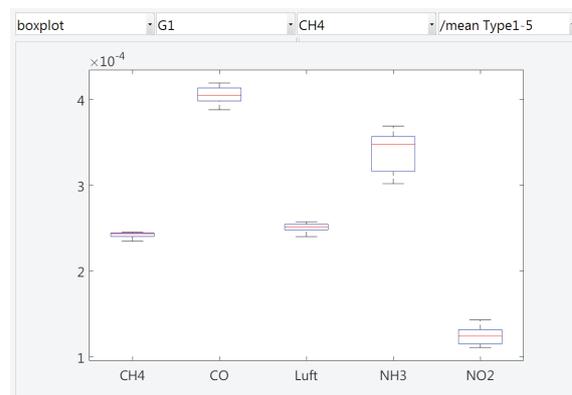


Abb. 6: Ansicht des Moduls „Visualization“, hier bei Auswahl des Boxplots. Dargestellt ist ein Mittelwert-Merkmal zur Unterscheidung fünf verschiedener Gase.

oder Balkendiagramm dargestellt werden. Die Korrelation aller Merkmalen untereinander resultiert hingegen in einer Matrix, die am sinnvollsten in einer sogenannten „Heatmap“ dargestellt wird, also einer zweidimensionalen Ansicht der Matrixwerte, in der hohe absolute Korrelation farblich besonders hervorgehoben wird. Entsprechende Merkmale enthalten ähnliche Information und sollten vor dem Training miteinander fusioniert oder alle bis auf eines entfernt werden [14].

Fazit und Ausblick

Wir haben eine umfangreiche MATLAB-Toolbox zur Auswertung zyklischer Sensordaten vorgestellt. Sie unterstützt den Anwender beim Finden von Strukturen in Daten beliebiger Sensoren, und führt ihn vom Import der Rohdaten bis zu einem Modell, das sich zur Klassifizierung neuer, unbekannter Daten eignet. Die Fusion verschiedener Sensoren geschieht dabei automatisch. Der grafische Aufsatz der Toolbox setzt den Fokus auf einfache Bedienbarkeit und ständige Visualisierung der Daten. Gleichzeitig können mit der kommandozeilenbasierten Version umfangreiche Skripte zur automatisierten Datenanalyse entwickelt werden.

In Zukunft soll die Toolbox noch deutlich erweitert werden, indem zum Beispiel auch Regressionsmodelle und entsprechende Zielvektoren zugelassen werden. Auch für Vorverarbeitung und Merkmalsberechnung sollen weitere Funktionen implementiert werden. Außerdem soll eine automatisierte Suche der für einen bestimmten Datensatz besten Vorverarbeitung und Merkmale realisiert werden, sowie eine Online-Auswertung, bei der Echtzeitmessdaten kontinuierlich eingelesen und anhand eines zuvor erstellten Modells klassifiziert werden.

Literaturnachweis

- [1] A.P. Lee, B.J. Reedy, *Sensors Actuators, B Chem.* 60, 35–42 (1999); doi:10.1016/S0925-4005(99)00241-5
- [2] P. Reimann, A. Schütze, *Gas Sensing Fundamentals*, Springer Berlin Heidelberg, Berlin, Heidelberg (2014); doi:10.1007/978-3-642-54519-1
- [3] U. Weimar, W. Göpel, *Sensors Actuators B Chem.* 26, 13–18 (1995); doi:10.1016/0925-4005(94)01547-U
- [4] N. Helwig, E. Pignanelli, A. Schütze, *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 210–215 (2015); doi: 10.1109/I2MTC.2015.7151267
- [5] C. Böhm, S. Berchtold, D. Keim, *ACM Comput. Surv.* 33, 322–373 (2001); doi:10.1145/502807.502809.
- [6] T. Baur, A. Schütze, T. Sauerwald, *Tm - Tech. Mess.* 82, 187–195 (2015); doi:10.1515/teme-2014-0007
- [7] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* 2 37–52 (1987); doi:10.1016/0169-7439(87)80084-9
- [8] R. Gutierrez-Osuna, *IEEE Sens. J.* 2, 189–202 (2002); doi:10.1109/JSEN.2002.800688
- [9] R. De Maesschalck, D. Jouan-Rimbaud, D.L.L. Massart, *Chemom. Intell. Lab. Syst.* 50, 1–18 (2000); doi:10.1016/S0169-7439(99)00047-7
- [10] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer (2009); doi:10.1007/b94608
- [11] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* 58, 109–130 (2001); doi:10.1016/S0169-7439(01)00155-1
- [12] A. J. Smola, B. Scholkopf, *Stat. Comput.* 14, 199–222 (2004). doi: 10.1023/B:Stco.0000035301.49549.88
- [13] HDF5 Group, HDF5, <https://www.hdfgroup.org/HDF5/> (accessed February 13, 2016)
- [14] T. Næs, B.H. Mevik, *J. Chemom.* 15, 413–426 (2001); doi:10.1002/cem.676