



1 *Labormitarbeiter testet die Benetzungseigenschaften eines Siliziumwafers.*

DIGITALE PRODUKTENTWICKLUNG BEI BASF

Eine der größten Herausforderungen in der Produktentwicklung in der chemischen Industrie ist es, mit möglichst geringem Aufwand eine kostengünstige chemische Zusammensetzung zu finden, welche gleichzeitig gewisse Qualitätseigenschaften hat.

Problemstellung, Projektbeispiel und Softwareentwicklung

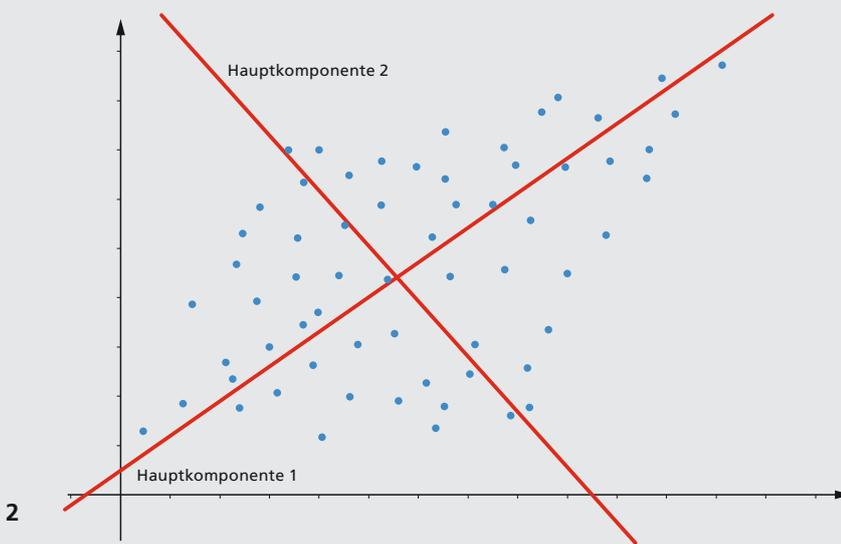
Ein Beispiel ist die Herstellung von Oberflächenbeschichtungen. Je nach Einsatzgebiet der Beschichtung sind verschiedene Eigenschaften wünschenswert. Neben dem Schutz der Oberfläche geht es beispielsweise auch um Glätte oder optische Eigenschaften. Nicht selten stehen sich die Ziele dabei gegenseitig im Weg, so dass ein geeigneter Kompromiss gefunden werden muss.

Ein typischer Workflow für ein neues Forschungsprojekt stellt sich wie folgt dar: Zunächst legen wir die Zielfunktionen und Designgrößen fest, bevor eine erste Versuchsreihe startet. Auf Basis der ersten Ergebnisse erstellen wir ein mathematisches Modell, um Vorhersagen für geeignete Designspezifikationen zu treffen. Darauf aufbauend wird eine neue Versuchsreihe gestartet und dieser Prozess solange iteriert fortgesetzt, bis eine zufriedenstellende Zusammensetzung gefunden ist.

Unsere Abteilung hat ein Softwaretool entwickelt, welches den Chemiker im gesamten Prozess begleitet. Das beginnt mit der Analyse und Visualisierung der Daten, setzt sich fort mit der Modellierung der einzelnen Zielgrößen bis hin zum Finden von optimalen Kompromissen und dem Planen von neuen Versuchen.

Maschinelles Lernen und Modellierung

Einen besonderen Stellenwert nimmt dabei die Wahl des mathematischen Modells ein, welches die Zielfunktionen in Abhängigkeit der Designvariablen beschreibt. Die Prozesse in der Chemie sind oft hochkompliziert und schwierig zu modellieren. In Zusammenarbeit mit der BASF SE entsteht gerade ein Tool, welches an dieser Stelle ansetzt und dabei gängige Algorithmen aus dem Maschinellen Lernen zum Einsatz bringt. Das bedeutet, dass alleine aus den bisherigen Daten ein Modell gewonnen wird, welches die Zusammenhänge beschreibt. Eine besonders komplexe Herausforderung ist die Wahl eines geeigneten Modells, denn dieses ist maßgebend für die Qualität der optimierten Lösungen.



Um die Tauglichkeit des Modelles zu messen, wenden wir verschiedene Verfahren an, wie zum Beispiel die Kreuzvalidierung. Sie nimmt nur einen Teil der Daten zum Training und den verbleibenden Teil zur Validierung des Modells. Genutzt wird sie zum Beispiel in der Komponentenauswahl für lineare Regressionsmodelle, um eine Überanpassung an die Daten zu vermeiden.

Konkret geht es darum, Komponenten mit geringem Einfluss, der lediglich zufälliger Natur ist, herauszufiltern und damit die Modellqualität zu verbessern. Allerdings führt ein rein datengetriebenes Modell nur in den wenigsten Fällen zum Ziel. Daher beziehen wir das Expertenwissen des Anwenders in den Modellierungsprozess mit ein.

Optimierung und Versuchsplanung

Ein zentraler Punkt ist die oben bereits angesprochene Planung neuer Versuche. Diese sind oft kostspielig und zeitintensiv; umso wichtiger ist eine effiziente Versuchsplanung. Das Tool unterstützt den Anwender dabei, mit möglichst wenigen Versuchen auszukommen und somit wertvolle Ressourcen zu sparen.

Gelöst wird dieses Problem wie folgt: Der Anwender selbst navigiert auf den Zielfunktionen in einen Bereich, welcher im Hinblick auf deren Werte besonders attraktiv ist. Voraussetzung dafür ist eine im Vorfeld durchgeführte mehrkriterielle Optimierung. In den Grenzen dieses Bereichs werden dann die neuen Versuche geplant. Auch die Vorwärtsplanung mithilfe des Modells unterstützt das Tool, sodass der Nutzer einzelne, besonders vielversprechende Rezepturen direkt testen kann. Der Anwender erhält dabei auch Informationen über die Unsicherheit der Vorhersage in Form von Konfidenzintervallen.

Web-Architektur in einem Big Data System

Realisiert haben wir die Software als Web-Lösung mit moderner Datenbank in einem Hadoop-System, welches eine Vielzahl von Datenmengen zu verwalten mag. Dies ermöglicht eine einfache und rechnerunabhängige Verwendung des Tools auch auf mobilen Endgeräten wie Tablets und Smartphones. Zugleich erhöht es die Produktivität in der Teamarbeit, weil alle Daten zentral gespeichert sind und Benutzer damit immer Zugriff auf den aktuellen Stand haben.

2 Die Abbildung zeigt die beiden Hauptkomponenten im zweidimensionalen Raum für einen Datensatz. Die Hauptkomponentenzerlegung teilt die ursprünglichen Größen durch Linearkombination in maximal aussagekräftige Größen und ordnet diese.

