

Formalisierung und Verwaltung von Entwicklungswissen im Kontext des Integrierten Produktmodells

Thilo Breitsprecher, Christoph Westphal, Nico Meintker und Sandro Wartzack

*Lehrstuhl für Konstruktionstechnik
Universität Erlangen-Nürnberg*

The structured management of product data has been one of the most important aspects in increasing the efficiency of development processes. But next to the access to product data the management of expert- and development knowledge is one key factor for successfully development. Therefore this paper should present an approach to acquire, store and to visualize knowledge by integrating a data mining process into an integrated product model.

1 Einleitung und wissenschaftliche Fragestellung

Die strukturierte Verwaltung von Produktdaten hat in den letzten Jahrzehnten maßgeblich zur Effizienzsteigerung in der Konstruktion aber auch in der Simulation beigetragen. Ausgehend von den EDM-Ansätzen der achtziger Jahre ist eine heterogene Systemlandschaft zur Verwaltung aller im Produktlebenszyklus anfallenden Daten entstanden, die auch in den nächsten Jahren zur Optimierung von Prozessen und Organisationen beitragen wird [1]. Dennoch ist für die Auslegung und Entwicklung von Produkten neben dem Zugriff auf bereits gespeicherte Daten vor allem der Zugriff auf existierendes Experten- und Entwicklungswissen von entscheidender Bedeutung. Doch hierzu fehlt es nach GERHARD nach wie vor an Konzepten das Expertenwissen des

Konstrukteurs „systematisch zu erfassen und so in eine formalisierbare Form zu bringen, dass es einerseits maschinenlesbar und trotzdem für den Konstrukteur verständlich und beherrschbar bleibt“ [7].

Aus diesem Gedanken lässt sich direkt die wissenschaftliche Fragestellung ableiten, in welcher Form Expertenwissen erfasst, formalisiert und abgebildet, also akquiriert werden kann (Knowledge Engineering). Diese Fragestellung ist jedoch nicht isoliert zu betrachten, sondern um den Aspekt des Managements von Konstruktions-, Simulations- bzw. Versuchsdaten zu erweitern. Neben der reinen Verwaltung von Daten und Wissen stellt sich darüber hinaus die Frage, wie dieses Wissen dem Entwickler in geeigneter Art und Weise zur Verfügung gestellt bzw. visualisiert werden kann.

Diese Fragestellung soll im Kontext des SFB Transregio 73 „Umformtechnische Herstellung von komplexen Funktionsbauteilen mit Nebenformelementen aus Feinblechen – Blechmassivumformung“ bearbeitet werden. Die im Rahmen dieses Beitrags entwickelten Methoden und Werkzeuge sollen es ermöglichen, dem Produktentwickler schnell und unkompliziert Auslegungsrichtlinien zur Verfügung zu stellen. Die angestrebte Visualisierung der Sensitivitäten und Abhängigkeiten zwischen Produkt- bzw. Prozessparametern einerseits und Produkteigenschaften andererseits soll die Qualität von Änderungskosten bei gleichzeitiger Reduzierung der Entwicklungszeit erhöhen.

2 Stand der Technik

2.1 Das TOTE-Schema als Problemlösungszyklus

Die wesentlichen Kerntätigkeiten im Entwicklungsprozess lassen sich nach WEBER auf Synthese – die Ableitung von Produktmerkmalen bzw. –parametern aus den Anforderungen – und Analyse – die Ermittlung der Produkteigenschaften anhand der festgelegten Produktmerkmale – reduzieren [14]. Eine Berücksichtigung der im Entwicklungsprozess üblichen Iterationen kann durch das TOTE-Schema nach MILLER realisiert werden [8]. Dabei wird ein Regelprozess beschrieben, der bei Nichterfüllung der Produkteigenschaften nach einer *Analyse (Test)* eine weitere *Synthese (Operate)* vorschlägt. Auf diese folgt eine weitere *Analyse (Test)*. Stimmen die geforderten Eigenschaften mit den ermittelten überein kann die Iterationsschleife beendet werden (*Exit*).

2.2 Verwaltung von Produktdaten

Dieser Problemlösungszyklus bedarf allerdings einer ganzheitlichen Verwaltung aller Daten über den gesamten Produktlebenszyklus, welche durch

Produktdatenmanagement-Systeme (PDMS) ermöglicht werden soll. Das Resultat einer gesamtheitlichen Modellierung aller Produktdaten wird als integriertes Produktmodell bezeichnet. Die bekannteste Form dieses Integrierten Produktmodells bietet die ISO 10303 bekannt als STEP [2], [6]. Sogenannte Application Protocolls stellen dabei eine anwendungsspezifische Instanz dieses Produktmodells zur Verfügung und dienen als datentechnische Grundlage für etliche kommerziell erhältliche PDM-Systeme wie Windchill PDMLink der Fa. PTC oder Teamcenter der Fa. Siemens UGS. Da diese Systeme ursprünglich zur Verwaltung der CAD-Daten konzipiert wurden, steht die Verwaltung der Konstruktionsdaten im Vordergrund.

Die Verwaltung der Simulationsdaten dient dagegen zur Unterstützung des Entwicklers bei der Dokumentation, Wiederverwendung und dem Vergleich Modelle aus aktuellen bzw. abgeschlossenen Projekten. Neben der ISO 10303 mit dem Application Protocoll AP209 *Composite and Metallic Structural Analysis and Related Design* ist vor allem der SimPDM-Ansatz zur Verwaltung von Simulationsdaten hervorzuheben [13]. Auch im Simulationsdatenmanagement sind etliche kommerzielle Systeme wie bspw. *SimManager* der Fa. MSC oder *Teamcenter for Simulation* der Fa. Siemens UGS verfügbar. Für die Verwaltung von physikalischen Messdaten bietet neben dem generischen STEP-Ansatz der ASAM ODS Standard eine wichtige Grundlage bei der Konzeption des Computer-Aided Testing [3].

2.3 Akquisition von Entwicklungswissen via Data Mining

Die genannten Ansätze verfolgen jedoch hauptsächlich die Verwaltung und Archivierung von Produktdaten. Zur Verwaltung von Entwicklungswissen sind neben der Wissenssicherung allerdings die Wissensakquisition und die Wissensbereitstellung notwendig.

Im Kontext des Knowledge Engineering nimmt die Wissensakquisition eine zentrale Rolle ein, ist aber gleichzeitig auch eine „Engstelle“, die den späteren Nutzen einer wissensbasierten Anwendung stark determiniert. In der Literatur sind hinsichtlich der Unterteilung der Wissensakquisition in Subprozesse unterschiedliche Tendenzen zu erkennen. Allen gemein ist aber der Standpunkt, dass die Wissensakquisition direkt, indirekt oder automatisiert erfolgen kann, wie in Bild 1 gezeigt [4], [5], [12].

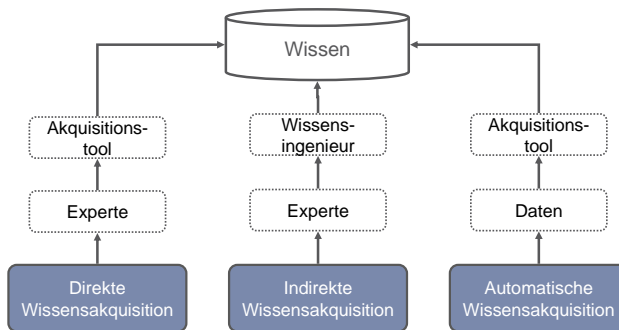


Bild 1: Methoden der Wissensakquisition nach [11]

Bezüglich formulierter Anforderungen an die Wissensakquisitionskomponente eines selbstlernenden Assistenzsystems zeigen die direkten und indirekten Methoden Defizite gegenüber automatisierter Vorgehensweisen. Daher wurde im des SFB/TR73 eine Methodik entwickelt, die es ermöglicht Entwicklungswissen aus Simulations- und Versuchsdaten automatisiert zu akquirieren [9], [10].

2.3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Kernprozess dieser Methodik ist die systematische Anwendung von Verfahren aus den Domänen der Statistik, der Künstlichen Intelligenz und des Maschinellen Lernens, die unter dem Sammelbegriff „Data Mining“ subsumiert werden [15].

Das CRISP-DM Referenzmodell sieht im Wesentlichen das Durchlaufen folgender Phasen vor: Das erste Teilziel ist ein *Aufgaben- und Prozessverständnis* (1) zu schaffen, zur Klärung der gesamten Data Mining Ziele und der gestellten Anforderungen aus Sicht des Anwenders. Es schließt sich die *Datensichtung und -selektion* (2) an, um ein grobes „Gefühl“ für die Daten zu entwickeln. Erste Hypothesen über vermutete Zusammenhänge können bereits aufgestellt werden. Die folgende *Datenvorverarbeitung und -transformation* (3) dient der Zusammenführung, Säuberung und gezielten Auswahl der Daten. Der erste Einsatz von Data Mining Methoden erfolgt in der Phase *Modellierung und Wissensentdeckung* (4) und liefert bereits eine Lösungen für das jeweilige Data Mining Problem. Die *Interpretation und Auswertung* (5) liefert Hinweise darauf, welche Entscheidung auf Basis der erreichten Ergebnisse zu treffen ist. Beispielsweise könnte ein zu schlechtes Prognosemodell die iterative Wiederholung der vorherigen Phasen notwendig machen. In der letzten Phase *Umsetzung* (6) sind die Ergebnisse so aufzube-

reiten, dass der adressierte Endnutzer sie direkt für seine Zwecke einsetzen kann.

2.3.2 Realisierung der Wissensakquisition

Die Grundlage für die Durchführung eines Data Mining Prozesses ist per se das Vorhandensein von Daten. Im Rahmen des SFB/TR73 fallen diese in Form von FE-Parametervariationen mit den entsprechenden Simulationsergebnissen sowie realen Umformversuchen an. Die „Achsen“ die diesen multidimensionalen Datenraum aufspannen sind „Produkt“ (z. B. ein bestimmtes Nebenformelement mit den Geometrie Größen - Produktmerkmale), „Prozess“ (Prozessmerkmale) und die erzielten „Ergebnisgrößen“ (z. B. der Umformgrad im Bauteil - Produkteigenschaften). Für eine spezifische Ausprägung dieser „Achsen“ steht nach einer Parametervariation oder einer Umformversuchsreihe ein Set an Daten für das Data Mining zur Verfügung. Mit Hilfe dieses Datensatzes sind die Algorithmen des Data-Minings in der Lage Vorhersagemodelle zu entwickeln, die in Abhängigkeit von Merkmalen (bspw. Geometrieparameter) festgelegte Eigenschaften (z. B. den Umformgrad) prognostizieren können [9], [15]. Dieses Schema ist durch Bild 2 wiedergegeben.

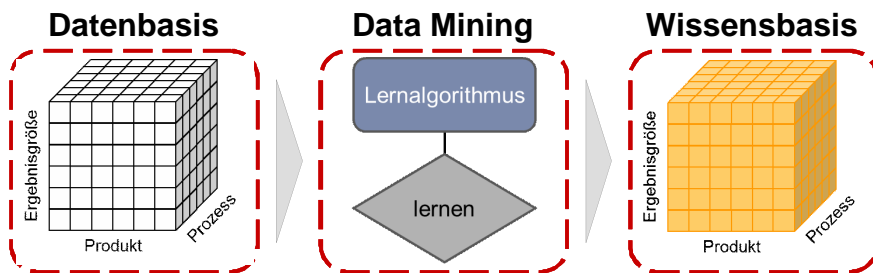


Bild 2: Prinzipieller Ablauf der automatischen Wissensakquisition via Data Mining

2.4 Ableitung des Handlungsbedarfs

Mit dem TOTE-Schema wurde ein etablierter Problemlösungszyklus vorgestellt, der den natürlichen Arbeitsablauf des Entwicklers darstellt. Um diesen effektiv zu unterstützen bedarf es nach Ansicht der Autoren Methoden das in der Analyse (Test) und der Synthese (Operate) benötigte Wissen effektiv in einem Produktdatenmanagementsystem zu verwalten und bereitzustellen.

Um die dazu notwendige systematischen Wissenserfassung und Formalisierung zu überwinden ist nach Ansicht der Autoren eine Integration des Data-Mining Prozesses in ein Produktdatenmanagementsystem zielführend.

3 Vorgehensweise für die Formalisierung und Verwaltung von Entwicklungswissen

Bei der Bereitstellung von Wissen muss dabei zwischen der Analyse und Synthese unterschieden werden. Einerseits kann die Analyse durch geeignete Vorhersagemodelle beschleunigt werden. Diese Vorhersagemodelle ermöglichen schnelle Aussagen zu den Produkteigenschaften ohne vorher zeitaufwändige Simulationen oder physische Versuche durchzuführen. Andererseits unterstützt eine Visualisierung der Sensitivitäten bzw. Korrelationen den Entwickler bei der Anpassung der Merkmale in der Synthese und vermeidet so unnötige Iterationsschleifen (Bild 3).

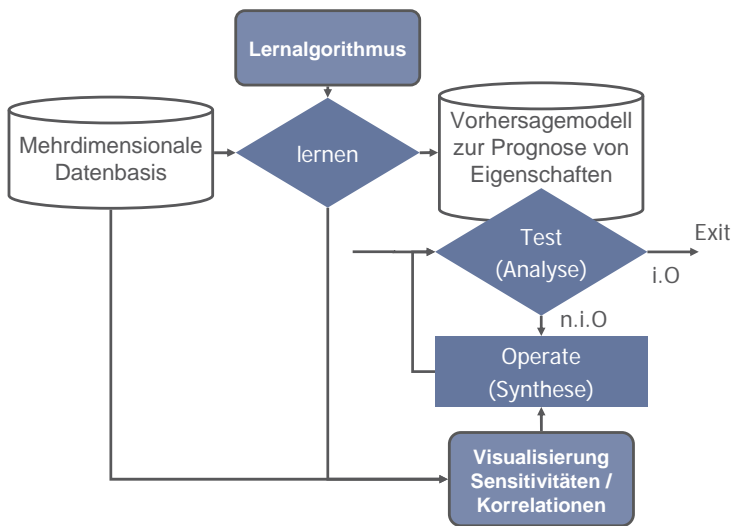


Bild 3: Angepasstes TOTE-Schema

Basis für die Erstellung des Vorhersagemodells bietet dabei die Verwaltung der Simulations- bzw. Versuchsparameter in einer Datenbank. Ein Data-Mining Tool (bspw. *RapidMiner* der i-Rapid GmbH) kann diese analysieren und verschiedene Arten von Vorhersagemodellen erzeugen. Die Auswahl der implementierten Vorhersagemodelle orientiert sich am vorliegenden Ziel des Teilprojektes: Lineare und Polynomiale Regression, Regellerner und Regres-

sionsbäume sind in der Lage aus numerischen Eingangsgrößen einer Datenbasis numerische Zielgrößen wie den Umformgrad oder auftretende Spannungen zu prognostizieren. Lernalgorithmen zur Prognose nomineller Zielgrößen (gut/schlecht; ja/nein; blau/grün/gelb/lila/usw.) können ebenfalls implementiert werden ohne vorher eine neue Datenbasis zur Verfügung stellen zu müssen. [15]

Aus diesem Prozess resultiert letztendlich ein Vorhersagemodell pro Produkteigenschaft (z.B. Umformgrad). Zusätzlich sind zu diesem Vorhersagemodell Fehlermaß und Standardabweichung bekannt. Dieses Vorhersagemodell kann nun auf Basis der aktuellen Konstruktionsparameter bzw. Produktmerkmale (z.B. Zahnhöhe) die zu erwartende Produkteigenschaft prognostizieren. Falls diese geforderte Produkteigenschaft nicht erfüllt werden, sind die Beitragsleister (Merkmale) für diese Eigenschaft zu ermitteln und dem Entwickler graphisch aufbereitet anzuzeigen. Die geänderte Merkmalskonfiguration kann anschließend durch das Vorhersagemodell wieder analysiert werden.

4 Implementierung und Vorhersage

Die Implementierung des Prozesses zur automatischen Wissensakquisition setzt sich aus einem Steuerprogramm mit grafischer Benutzerschnittstelle, einem Prozess-Steuerprogramm, einem relationalen Datenbanksystem sowie einem System für die Durchführung des Data Minings zusammen.

4.1 Implementierung der Datenbankstruktur

Grundlage für ein effizientes Datenmanagement ist ein Produktmodell. Dieses orientiert sich am STEP-Produktmodell, besteht aus 24 Entitäten und ist im relationalen Datenbankverwaltungssystem MySQL 5 implementiert. Somit können sämtliche relevanten Informationen eines Umformvorganges sowie die daraus resultierenden Eigenschaften aufgenommen werden. Relevante Informationen sind u.a. die Parameter der verwendeten Konzepte für Werk- und Halbzeug, spezifische Fertigungsverfahren sowie Geometriegrößen des Bauteils mit seinen Haupt- und Nebenformelementen (Merkmale). In diesem Aspekt spiegelt sich auch der in 2.3.2 beschriebene, multidimensionale Charakter des Datenraumes wider. Weitere Funktionalitäten sind die Zuordnung eines Vorhersagemodells zu einem Bauteil und die Archivierung älterer Modelle.

4.2 Realisierung des Selbstlernprozesses

Über eine graphische Oberfläche werden dem Anwender nur diejenigen Bauteile und Eigenschaften zur Verfügung gestellt, zu denen eine Vorhersage erstellt werden kann. Von dieser GUI aus kann der Nutzer anschließend den Selbstlernprozess anstoßen. Der nun ausgeführte Programmcode sorgt für die Steuerung des Data Mining Prozesses sowie für die notwendigen Datenbankoperationen zum Aus- und Einlesen aller benötigten und anfallenden Daten. Der gesamte Vorgang ist zusammenfassend in Bild 4 dargestellt und wird nachfolgend im Detail erklärt.

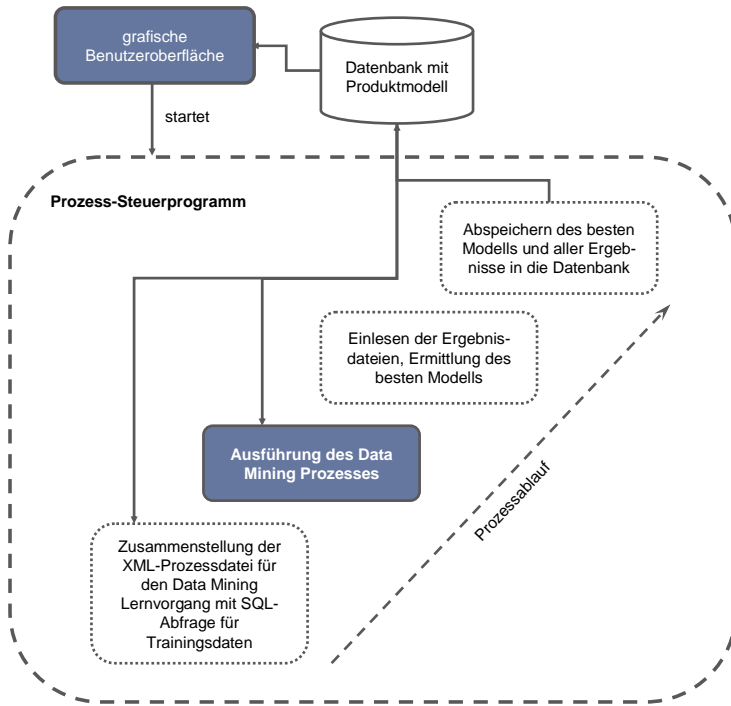


Bild 4: Ablauf des Prozesses zur automatischen Wissensakquisition mit Datenbankinteraktion

Nach Anstoß des Selbstlernprozesses durch den Anwender erfolgt zunächst die Erstellung eines Templates, welches den Data Mining Prozess vorgibt. In diesem Template erfolgt die Festlegung des korrekten Ausgabeverzeichnis für die Ausgabedateien aus dem Data Mining Prozess. Anschließend erfolgt die Erstellung und Import des SQL-Strings zur Auswahl der Trai-

ningsdaten aus der Datenbank. Nachdem das Template des Data Mining Prozesses erstellt ist, wird der Data Mining Prozesses in dem Data Mining System *RapidMiner* in der Konsolenausgabe gestartet, so dass der Prozess zur Erstellung der Modelle abgearbeitet wird. Nach Abschluss des Data Mining Prozesses erfolgt das Auslesen aller Modelle und Performances sowie der gepaarten T-Tests aus den Ausgabedateien. Dabei werden die Daten aus den Ausgabedateien in den Arbeitsspeicher zur weiteren Verarbeitung übertragen. Bei der Auswertung werden zunächst alle Modelle gesucht, die eine signifikant geringere Vorhersagegenauigkeit aufweisen als irgendein anderes Modell. In diesem Falle scheidet das unterlegene Modell aus der finalen Auswahl aus. Unter den Modellen, welche sich noch in der finalen Auswahl befinden wird dasjenige als bestes Modell gewählt, welches das geringste Fehlermaß aufweist. Falls zwei Modelle das gleiche geringste Fehlermaß besitzen wird dasjenige mit der geringsten Streuung gewählt. Im seltenen Fall, dass mehrere Modelle sowohl das gleiche geringste Fehlermaß als auch die gleiche Streuung besetzen, wird das zuerst gefundene Modell als bestes Modell markiert. Abschließend wird ein SQL-String erstellt und ausgeführt, welcher die Aufgabe hat alle Ergebnisse aus dem Data Mining Prozess in die Datenbank gemäß der Struktur des Produktmodells zu übertragen. Zudem wird auch das beste ermittelte Modell abgespeichert.

4.3 Der Data Mining Prozess

Aufgabe des Data Mining Prozesses ist es aus Fertigungs- und Simulationsdaten des Umformprozesses automatisch ein bestmögliches Vorhersagemodell für ein bestimmtes Bauteil und für eine vorgegebene Eigenschaft zu erstellen. Hierzu wurde ein 127 Operatoren umfassender Data Mining Prozess erstellt, welcher die gesamte Wissensakquisition ausführt. Dieser Prozess soll im nachfolgenden erklärt werden. Nach dem Auslesen der Daten aus der Datenbank werden zunächst alle numerischen Werte als solches markiert, die Ergebnisgröße wird als Zielfunktion festgelegt. Zudem werden alle Tupel mit fehlender Zielfunktion entfernt. Parameter die sich nicht ändern und dementsprechend eine Standardabweichung von null aufweisen werden ebenfalls entfernt. Um das geforderte, bestmögliche Vorhersageergebnis zu erhalten werden in dem Data Mining Prozess neben den Standardeinstellungen der vier Lernfunktionen - Lineare und Polynomiale Regression, Regellerner und Regressionsbäume - für jede Lernfunktion zusätzlich sowohl eine schrittweise als auch eine evolutionäre Parameteroptimierung durchgeführt. Als weitere Maßnahme wird versucht durch einen „Forward Selection“-Operator, bei der verbrauchte Parameter entfernt werden, das Ergebnis zusätzlich zu verbessern. Somit ergeben sich pro Lernfunktion insgesamt sechs verschiedene Modelle, so dass bei vier Lernfunktionen insgesamt 24 Modelle entstehen. Zu den Mo-

dellen wird per 10-facher Kreuzvalidierung das Fehlermaß ermittelt und Anschließend ein T-Test durchgeführt. Alle Ausgaben der Performances, Modelle und T-Tests werden abschließend in den jeweiligen Ausgabedateien abgespeichert. Nachfolgend ist eine vereinfachte Version der grafischen Darstellung des Data Mining Prozesses dargestellt, welcher nur mit einer anstatt vier Lernfunktionen ausgeführt ist (Bild 5).

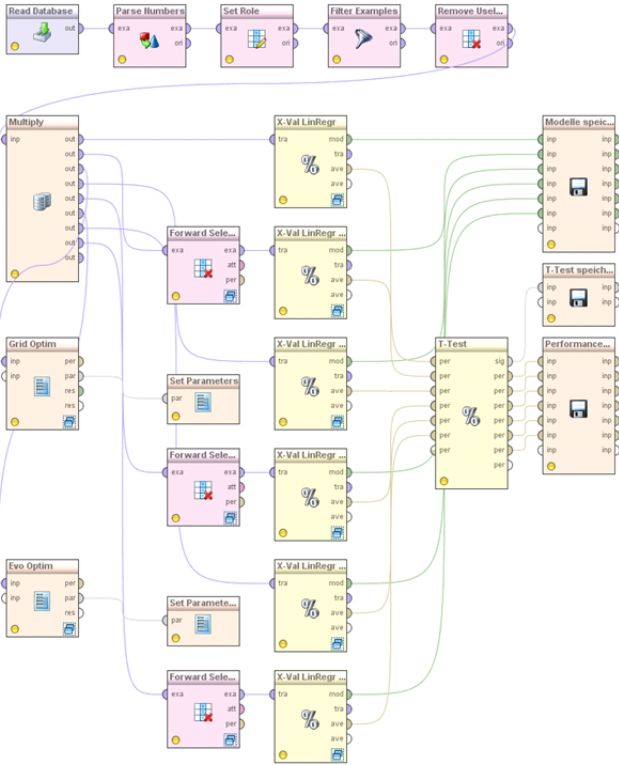


Bild 5: Darstellung eines vereinfachten Data Mining Prozesses mit nur einer Lernfunktion

5 Zusammenfassung, Diskussion und Ausblick

In dieser Arbeit wurde eine Methodik vorgestellt, wie der natürliche Problemlösungszyklus des Produktentwicklers mit Vorhersagemodellen unterstützt werden kann, die in einer automatischen Wissensakquisition per Data Mining erstellt und in einer Datenbank verwaltet werden können. Somit kann die Analyse der Produkteigenschaften beschleunigt und die Dauer einer Iterati-

onsschleife reduziert werden. Diese Methodik eignet sich besonders bei Änderungskonstruktionen, bei denen auf eine bestehende Datenbasis von Vorgängerprodukten zugegriffen werden kann.

Ein Vorhersagemodell beschreibt die Korrelation zwischen einer Eigenschaft und den zugrundeliegenden Merkmalen. Im Syntheseprozess wird allerdings vor allem die Korrelation zwischen einem Merkmal und den daraus resultierenden Eigenschaften benötigt. Daher steht die Ableitung der Sensitivitäten bzw. Korrelationen zwischen Merkmalen und Eigenschaften aus der Datenbasis aber auch vor allem aus den Vorhersagemodellen im aktuellen und zukünftigen Fokus der Arbeiten. Gelingt dies kann auf Basis der Vorhersagemodelle eine ganzheitliche Unterstützung der Synthese und Analyse in der Produktentwicklung geleistet werden.

Danksagung

Die Autoren möchten an dieser Stelle der Deutschen Forschungsgemeinschaft (DFG) danken, die den SFB Transregio 73 fördert.

Literatur

- [1] Abramovici, M., Schindler, T.: "Product Lifecycle Management (PLM), in: "15. Newsletter Berliner Kreis", Paderborn, 2010.
- [2] Anderl, R., Trippner, D.: "STEP Standard for the Exchange of Product Model Data. Eine Einführung in die Implementierung und industrielle Nutzung der Normenreihe ISO 10303", B. G. Teubner, Stuttgart, 2000.
- [3] Bartz, R.: "Association for Standardization of Automation and Measuring Systems", Hoehenkirchen, 2006.
- [4] Diaper, D.: "Knowledge elicitation. Principles, techniques and applications", New York, 1989.
- [5] Frick, D.: „Die Akquisition betriebswirtschaftlichen Wissens zum Aufbau von wissensbasierten Entscheidungsunterstützungssystemen“. Universität Duisburg, Dissertation, 1997.
- [6] ISO. DIN EN ISO 10303 AP 214 Standard for the exchange of product model data. 1985.

-
- [7] Gerhard, D., Lutz, C.: "Rechnergestütztes Konfigurieren und Auslegen kundenindividueller Produkte, in: "15. Newsletter Berliner Kreis", Paderborn, 2010.
- [8] Miller, G.A., Galanter, E., Pribram, K.: „Plans and the structure of behaviour“, New York, 1960.
- [9] Röhner, S.; Breitsprecher, T.; Wartzack, S.: „Anforderungen an die Wissensakquisitionskomponente eines selbstlernenden Assistenzsystems“, in: „8. Gemeinsamen KT-Kolloquium“, Magdeburg, 2010.
- [10] Röhner, S.; Breitsprecher, T.; Wartzack, S.: "Acquisition of Design-relevant Knowledge within the Development of Sheet-Bulk metal forming", in: "International Conference on Engineering Design", Kopenhagen, 2011.
- [11] Spur, G.; Krause, F.-L.: „Das virtuelle Produkt. Management der CAD-Technik“, München, 1997.
- [12] Puppe, F.: „Problemlösungsmethoden in Expertensystemen“, Berlin, 1990.
- [13] PROSTEP IVIP, E. V. Integration der Simulation und Berechnung in eine PDM-Umgebung. Darmstadt. 2008.
- [14] Weber, C.: „CPM/PDD - An extended theoretical approach to modelling products and product development process“, in Proceedings of 2nd German-Israeli Symposium on Advances in Methods and Systems for Development of Products and Processes, Fraunhofer-IRB-Verlag Berlin, 2005.
- [15] Witten, I.: "Data Mining – Practical Machine Learning Tools and Techniques", Burlington, 2011.