

Information zum Referenten:

Dr. Thomas Natschläger

**Head of Industrial Data Mining Group
Software Competence Center Hagenberg**



Studium

Technische Universität Graz, Austria

1989 – 1995 Studium Telematik mit Schwerpunkt Theoretische Informatik

1999 Promotion mit einem Thema aus dem Bereich Computational Neuroscience

Berufliche Karriere

Apr. 1996 – Sept. 1999: Wissenschaftlicher Mitarbeiter am Institut für Grundlagen der Informationsverarbeitung, Technische Universität Graz: Forschung im Bereich Computational Neuroscience und Machine Learning

Sept. 1999 – Sept. 2003: Universitätsassistent und Lektor am Institut für Grundlagen der Informationsverarbeitung, Technische Universität Graz: Forschung und Lehre im Bereich Computational Neuroscience und Machine Learning

Derzeitiger Arbeitgeber / Institution

Seit Oktober 2003 Software Competence Center Hagenberg (SCCH), Österreich

Projektmanagement

Leitung der Gruppe Industrial Data Mining

www.scch.at

thomas.natschlaeger@scch.at

Wissensextraktion mittels maschinellen Lernens zum besseren Prozessverständnis in der Stahlerzeugung

*Thomas Natschläger¹, Johannes Himmelbauer¹, Holger Schöner¹,
Friedrich Kokert², Silvia Mitter², Herbert Exner³*

¹ *Software Competence Center Hagenberg, SCCH*

² *voestalpine Stahl GmbH*

³ *uni software plus GmbH*

Zusammenfassung

Die intelligente Analyse von Produktionsdaten beinhaltet ein großes Potential zur Extraktion von neuem Wissen über den Produktionsprozess. Dieses aus den Daten generierte Wissen kann genutzt werden, um Optimierungspotentiale zu orten, die in weiterer Folge zur Prozessverbesserung führen können. In dieser Arbeit soll anhand eines Beispiels aus der Stahlerzeugung in der voestalpine Stahl gezeigt werden, wie ein solcher Wissensextraktionsprozess in einer Kooperation zwischen Datenanalyseexperten und Prozessexperten umgesetzt werden kann. Im speziellen konzentrieren wir uns auf den Prozessschritt des Stranggießens. Dabei liegt das Hauptaugenmerk auf der Fragestellung, welche Faktoren die Vergießbarkeit des Stahles an den Stranggussanlagen beeinflussen. Wir beschreiben in diesem Beitrag eine Analyseumgebung, die speziell auf diese Fragestellung abgestimmt ist und es zukünftig den Experten in der voestalpine Stahl GmbH erlauben wird, den Wissensextraktionsprozess für neue Produktionsdaten einfach und effizient durchführen zu können.

Keywords: Wissensextraktion, Prozessverständnis, Automatisierung von Wissensextraktion, Maschinelles Lernen

1 Einleitung

In einer langfristigen Kooperation zwischen der Software Competence Center Hagenberg GmbH (SCCH) und der voestalpine Stahl GmbH werden verschiedene Prozessschritte in der Stahlproduktion ausgewählt, und in enger Zusammenarbeit mit den Experten der voestalpine Stahl GmbH wird versucht, das in den dafür verfügbaren Produktionsdaten vorhandene Wissen bzw. Optimierungspotential zu extrahieren und in weiterer Folge zur Prozessverbesserung zu nützen.

Einer dieser Produktionsschritte liegt an der Stranggussanlage, der Ort, wo der flüssige Stahl aus einem Verteiler durch ein Gießrohr zu Brammen vergossen wird (siehe Abbildung 1).

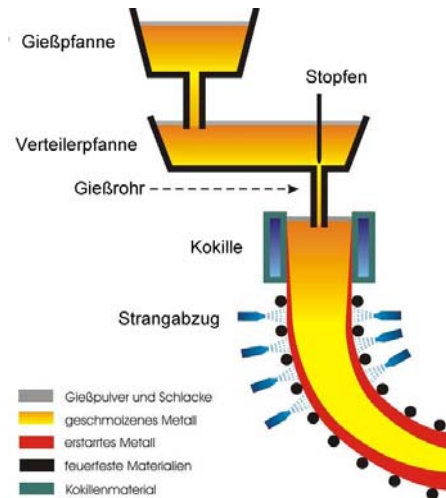


Abbildung 1: Links: Schematische Darstellung einer Stranggussanlage. Rechts: Blick in die Kokille einer Stranggussanlage der voestalpine Stahl GmbH (Quelle: www.expeditionvoestalpine.com)

An dieser Stelle tritt das Problem des sogenannten “Clogging” [Trueba et al. 2006] auf, d.h. der Durchfluss des Stahls verschlechtert sich mit der Zeit aufgrund von Ablagerungen im Gießrohr und am Stopfen, der den Durchfluss ins Gießrohr regelt. Dieses Phänomen des Clogging macht es notwendig, Verteiler und Gießrohr in bestimmten Abständen zu wechseln. Eine Verbesserung der Vergießbarkeit und eine damit verbundene, mögliche Verlängerung der Gießsequenzen stellen ein großes Einsparungspotential dar.

Die hohe Komplexität dieses Ablagerungsprozesses mit seiner Vielzahl von möglichen Einflussgrößen (sowohl physikalische wie z.B. die Gießgeschwindigkeit oder die Anlagenkonfiguration als auch die chemische Zusammensetzung des Stahls) und den sehr vielen verschiedenen Szenarien (unterschiedliche Stahlsorten mit unterschiedlichen Vorbehandlungen und vier Stranggussanlagen mit unterschiedlichen Anlagenkonfigurationen) erschweren die analytische Modellierung des Clogging-Problems.

Die gestellte Aufgabe in der aktuellen Kooperation ist es, die wichtigsten Einflussgrößen bzw. Kombinationen von diesen aus einer Vielzahl von möglichen Kandidaten (~ 150) für viele verschiedene Szenarien (~ 100) möglichst automatisiert zu bestimmen, zu charakterisieren und zu quantifizieren. Dabei kommen verschiedenste Methoden des maschinellen Lernens zum Einsatz. Die auf diese Weise erhaltenen Ergebnisse sollen für die Stahlexperten der Forschungsabteilung der voestalpine Stahl GmbH eine gute Basis bilden und hilfreiche Hinweise auf Einsparungsmöglichkeiten aufzeigen, die dann nach Überprüfung auf Durchführbarkeit im Stahlwerk umgesetzt werden sollen.

Die dafür notwendigen Arbeiten lassen sich grob in drei Abschnitte unterteilen, die in den folgenden Abschnitten detaillierter beschrieben werden. Die erste Aufgabe bestand in der Datenaufbereitung (siehe Abschnitt 2). Neben der Datenbereitstellung aus verschiedenen Datenbanken und der geeigneten Aufarbeitung bzw. Generierung von möglichen Einflussgrößen (Abschnitt 2.1) bestand hier die Herausforderung insbesondere darin, ein geeignetes Maß für die nicht direkt messbare Zielgröße, die Vergießbarkeit, zu erarbeiten (siehe Abschnitt 2.2). In Abschnitt 3 wird eine Analyseumgebung beschrieben, mit der unter möglichst einfacher Handhabung Untersuchungen nach oben angeführten Anforderungen durchgeführt und die erzielten Ergebnisse in geeigneter Weise zusammengefasst und präsentiert werden können. Dabei kommen verschiedenste Algorithmen des maschinellen Lernens zur Wissensextraktion zur Anwendung. In der dritten Phase (Abschnitt 4) wird

das so geschaffene Analysewerkzeug den Bedürfnissen der Stahlexperten entsprechend erweitert und adaptiert, wobei das Ziel verfolgt wird, dass der Wissensextraktionsprozess direkt von den Experten in der voestalpine Stahl GmbH auf neue Produktionsdaten angewendet werden kann.

2 Datenaufbereitung

2.1 Datenintegration und Merkmalsextraktion

Die zur Untersuchung zur Verfügung gestellten Daten bestehen grundsätzlich aus zwei Teilen; zum einen aus den Stranggussdaten, die die Produktionsdaten des Gießvorgangs an den Stranggussanlagen beinhalten (Gießgeschwindigkeit, Kokillenbreite, Verteilernummer, Stopfenstellung, Füllgrad im Verteiler, etc.), und zum anderen aus Tabellen mit Daten der Sekundär-Metallurgie, die vor allem Informationen über die chemische Zusammensetzung und die Behandlungsvorgeschichte der Schmelzen liefern (hauptsächlich Analysen, Proben, Behandlungen, Behandlungszeiten, Materialzugaben, -zeiten, und -orte). Die Daten liegen dabei in unterschiedlicher Granularität vor: Die Daten der Sekundär-Metallurgie liegen pro Schmelze vor. Die Stranggussdaten werden je Segment einer vergossenen Bramme aufgezeichnet. Es ist daher notwendig, die Stranggussdaten so zu aggregieren, dass sie eindeutig mit den Daten der Sekundär-Metallurgie verknüpft werden können (über eine eindeutige Schmelzennummer), um die Vergießbarkeit je Schmelze zu bestimmen bzw. untersuchen zu können. Nach Generierung von weiteren, abgeleiteten Merkmalen aus den schon vorhandenen erhält man einen Datensatz mit ca. 270 Merkmalen, welche mögliche Einflüsse auf die Vergießbarkeit darstellen.

2.2 Definition eines Maßes für die Vergießbarkeit

Das Phänomen des Cloggings, welches die Vergießbarkeit negativ beeinflusst, ist nicht direkt messbar, sondern kann nur indirekt über den Stopfenverlauf beobachtet werden. Der Stopfen regelt den Durchfluss des flüssigen Stahls ins Gießrohr. Wenn sich nun am Stopfen (oder im Gießrohr selbst) Ablagerungen bilden, muss der Stopfen erhöht werden, um eine gleichbleibende Gießleistung zu gewährleisten (siehe Abbildung 2, rote Linie).

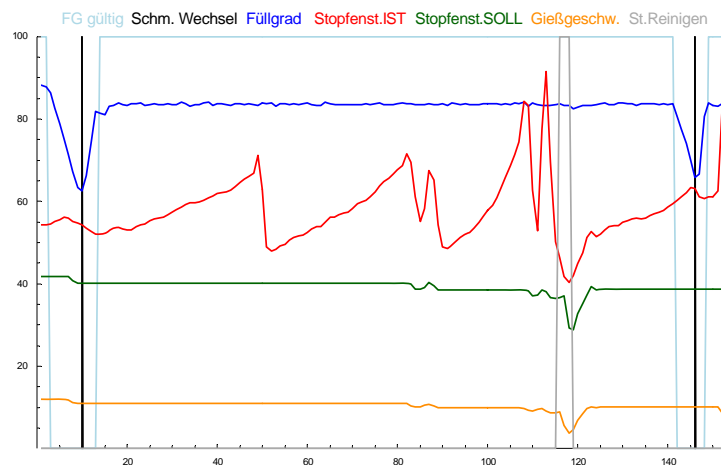


Abbildung 2: Darstellung eines typischen Verlaufs der Stopfenstellung (rote Linie).

Grundidee bei der Definition eines objektiven Maßes für die Vergießbarkeit ist es, diese Folgeerscheinung des Cloggings zu nützen und je Schmelze zu quantifizieren. Dabei kommen jedoch Effekte zum Tragen, die eine genaue Bestimmung der Vergießbarkeit erschweren. Hauptproblem ist,

dass nicht jede Änderung der Stopfenstellung eine Folge der Ablagerungen sein muss: Eine Änderung der Gießgeschwindigkeit (orange Linie) wird zum Beispiel durch eine Erhöhung bzw. Senkung des Stopfens erzielt. Wird erkannt, dass starkes Clogging eingetreten ist, wird oft der Stopfen „durchgedrückt“ (d.h. der Stopfen stark nach unten gesenkt), um eventuell ein Abbrechen der schon vorhandenen Ablagerungen zu erzielen. In Abbildung 2 sieht man dies im Stopfenverlauf an den starken Sprüngen nach unten. Bei der Entwicklung eines geeigneten, objektiven Bewertungskriteriums zur Bestimmung der Vergießbarkeit musste also versucht werden, möglichst nur Stopfenerhöhungen, die Folge von Ablagerungen waren, bei der Berechnung zu berücksichtigen. Ein weiteres Problem bei der Findung eines geeigneten Maßes waren die unterschiedlichen Arten des Cloggings (Stopfen-, bzw. Gießrohrschmier). Zum einen können diese unterschiedliche Ursachen haben, und zum anderen spiegelt sich das Gießrohr-Clogging in der Regel nicht sofort zu Beginn im Stopfenverlauf wider. Erst wenn der Durchflussdurchmesser im Gießrohr durch Ablagerungen zu klein wird, erfolgt plötzlich ein rasanter Anstieg der Stopfenstellung. Dies kann zur Folge haben, dass von einer Schmelze A verursachte Ablagerungen erst bei einer späteren Schmelze B zum Tragen kommen und das Auftreten des Cloggings fälschlicherweise alleinig der Schmelze B zugeschrieben wird. Aus diesen Gründen hat man sich dazu entschieden, den errechneten „allgemeinen Clogging-Index“ um das Gießrohrschmier (erkennbar durch den dadurch verursachten, exponentiellen Stopfenverlauf) zu bereinigen.

Ergebnis dieser Arbeit war je ein eigenes, neues Maß für das Stopfen-Clogging („SCI“) und das Gießrohr-Clogging („GCI“). Es wurden damit Kennzahlen geschaffen, die nach Einschätzung der Stahl-Experten das Potential in sich tragen, die bisher verwendeten Größen für das Phänomen der Vergießbarkeit besser und verlässlicher beschreiben zu können und die somit nach weiteren Tests und eventuellen Verbesserungen als generelle Zielgrößen für Vergießbarkeitsanalysen eingeführt werden sollen.

2.3 Datenbereinigung

Ein entscheidender Punkt bei der Datenaufbereitung war die Datenbereinigung und die damit verbundene Datenüberprüfung. Einerseits ist die Datenbereinigung notwendig, um repräsentative Einflüsse bzw. Zusammenhänge aus den Daten zu extrahieren. Ausreißer und unplausible, fehlerhafte Werte in den Eingangsgrößen können eigentlich vorhandene Zusammenhänge mit der Vergießbarkeit verschleiern bzw. Vorhersagemodelle für die Vergießbarkeit qualitativ stark beeinträchtigen. Andererseits gibt die Datenüberprüfung die Möglichkeit, Fehler bzw. Ungereimtheiten in den vorhandenen Datenbanken zu erkennen, und diese wertvollen Erkenntnisse können wiederum zur Erhöhung der Datenqualität in den einzelnen Datenbeständen genutzt werden.

Das angemessene Filtern der Daten nach Ausreißern stellte sich in dieser Anwendung als sehr schwierig heraus. Ein Hauptziel ist es, in den Daten Hinweise zu finden, die zu Ursachen von besonders schlechter Vergießbarkeit führen. Ausnahmesituationen, d.h. Anomalien in den Eingangsgrößen, können diesbezüglich besonders interessant sein. Außerdem war ein umfassendes, händisches Überprüfen der Merkmalsverteilungen aufgrund der Vielzahl an Merkmalen und Szenarien unmöglich. Letztendlich wurde ein automatischer Filtermechanismus, basierend auf Quantilwerten, eingesetzt. Eine geeignete, visuelle Darstellung der potentiellen Einflüsse bildet die Grundlage, für einzelne Szenarien und Merkmale die Filterung vor weiteren Analysen händisch zu korrigieren. Die Schwierigkeiten bei der Ausreißerbehandlung waren auch eine zusätzliche Motivation, Fehler bzw. Ungereimtheiten in den Daten bei der Datenaufbereitung zu finden und möglichst an der Wurzel zu beheben, um so schon vor der Filterung einen konsistenten Datensatz von möglichst hoher Qualität zu erhalten.

3 Wissensextraktion mittels maschinellem Lernen

Die im obigen Abschnitt beschriebene Datenaufbereitung stellt den Ausgangspunkt – die aufbereiteten Produktionsdaten - für den eigentlichen Analyseprozess dar. Der allgemeine Prozess der Wissensextraktion ist in Abbildung 3 dargestellt.

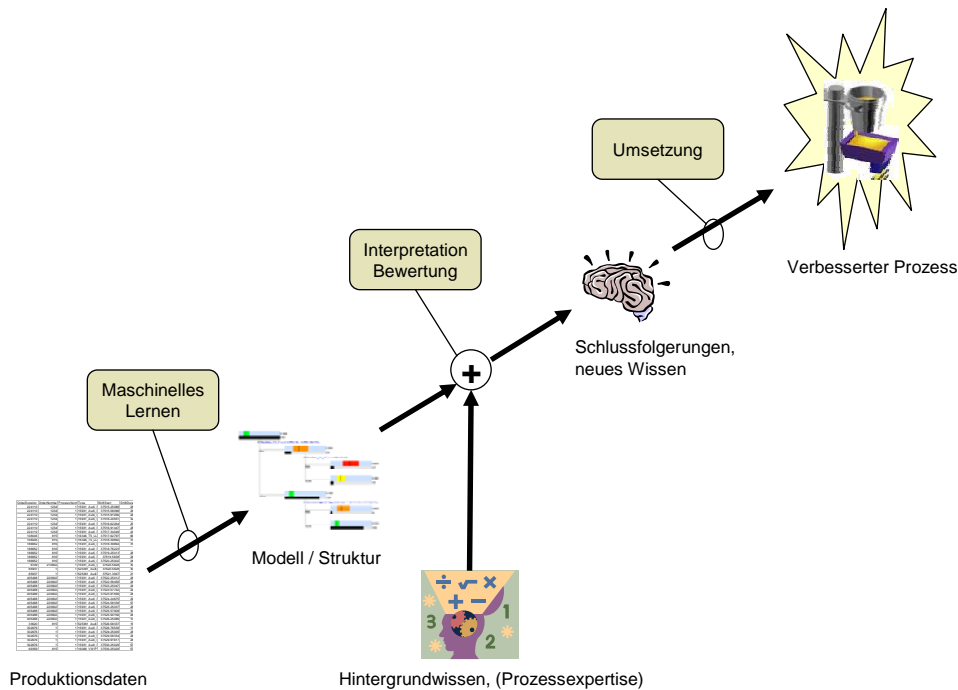


Abbildung 3: Darstellung des Wissensextraktionsprozesses basierend auf maschinellem Lernen.

Dabei werden aus den Produktionsdaten mit Hilfe von Methoden aus dem Bereich des maschinellen Lernens [Duda et al. 2001] und der Statistik [Hastie et al. 2001] Modelle generiert, welche in strukturierter Weise Zusammenhänge zwischen potentiellen Einflüssen und dem relevanten Zielmerkmal (hier der Clogging-Index) beschreiben. Diese unbewertete, strukturelle Information muss nun mit dem Wissen der Prozessexperten interpretiert und bewertet werden. Erst durch diesen Vorgang können Schlussfolgerungen, welche neues Wissen über den Produktionsprozess bedeuten, formuliert werden. Die Umsetzung vielversprechender Schlussfolgerungen erlaubt letztendlich eine Verbesserung des Produktionsprozesses.

3.1 Verwendete Methoden des maschinellen Lernens

Die Prozessexpertise ist in erster Linie als know-how in den Köpfen der Prozessexperten vorhanden. Daher ist es derzeit nicht möglich, den Vorgang der Interpretation und Bewertung der generierten Modelle zu automatisieren. Es ist jedoch von immensm Nutzen, dass die generierten Modelle Strukturen aufweisen, welche mit jenen möglichst kompatibel und vergleichbar sind, in denen die Prozessexpertise vorliegt. In der konkreten Anwendung wurde daher großes Augenmerk darauf gelegt, dass die verwendeten Methoden Modelle erzeugen, die bereits auf Grund ihrer Struktur gut von den Prozessexperten interpretierbar sind. Zusätzliche Visualisierungen ergänzen diese Modelle und erlauben so schlussendlich eine effiziente Bewertung und Interpretation der Produktionsdaten.

Eine Gruppe von Methoden, die für diesen Kontext gut geeignet sind, stellen Fuzzy-Logic-basierte Regeln bzw. Entscheidungsbäume dar, da die einfache Regelstruktur dem menschlichen Denken sehr entgegenkommt. In Abbildung 4 (links) ist ein solcher, Fuzzy-Logic-basierter Entscheidungsbaum

dargestellt, welcher durch den LIRT-Algorithmus [Drobics, Himmelbauer 2007, Drobics 2005] generiert wurde.

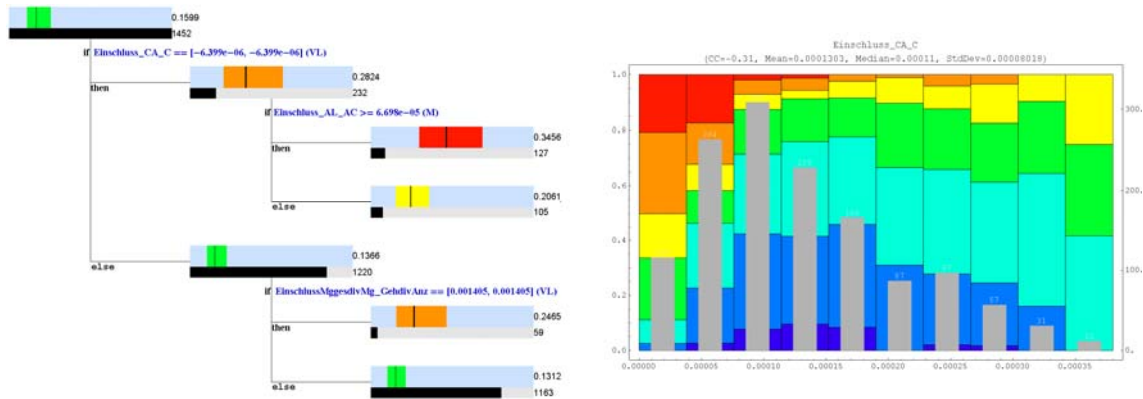


Abbildung 4: Links: Darstellung eines Entscheidungsbaumes. Rechts: Visualisierung des Histogramms einer möglichen Einflussgröße. Erstellt mit dem machine learning framework for Mathematica.

Dieser Entscheidungsbaum beschreibt mit einfachen *if-then-else-Regeln* mögliche, qualitative Gründe für schlechte (roter Balken) und für gute (grüner Balken) Vergießbarkeit und stellen einen sehr praktikablen Ausgangspunkt für eine Bewertung dar. Um diese Bewertung weiter zu vereinfachen, werden in dieser konkreten Anwendung übersichtliche Datenvisualisierungen den Modellen beigelegt. Ein Beispiel dafür ist in Abbildung 4 (rechts) dargestellt. Diese Art von Histogramm mit farblich hinterlegter Güte der Vergießbarkeit erlaubt es, auf visueller Basis die potentiellen Einflüsse rasch qualitativ zu bewerten.

Da es für eine gegebene Aufgabenstellung nicht klar ist, welche Methode des maschinellen Lernens zu den besten Modellen führt, werden auch in dieser Anwendung mehrere Modelle parallel betrachtet. Insbesondere wird neben der klassischen, multivariaten linearen Regression eine Regressionsmethode angewandt, welche es explizit erlaubt zu untersuchen, welche Kombinationen / Interaktionen von Einflüssen relevant sein könnten. Diese sogenannte iterative quadratische Regression [Natschläger 2008] erstellt ähnlich zur linearen Regression eine Linearkombination von möglichen Einflussgrößen. Im Gegensatz zur linearen Regression können die einzelnen Terme jedoch auch Produkte von Variablen enthalten. In Abbildung 5 ist eine Visualisierung eines solchen Modells dargestellt.

| Weight | Attribute |
|----------|--|
| 1.552 | |
| 164.4 | $(-0.1503 + 0586 \text{ CC.GIRO.GR-NUT-DRCK Rep. Meßwert}) (-0.004596 + \text{EinschlussMggesdivMg_GehdivAnz})$ |
| -0.03214 | 0631 CC.STRG.Gießleistung Rep. Meßwert |
| -1.337 | EinschlussMgdivAl_Geh |
| 298.8 | Einschluss_AL_AM |
| 1.02 | Einschluss_MG_Z |
| -0.1737 | 0586 CC.GIRO.GR-NUT-DRCK Rep. Meßwert |
| 0.07036 | $(-2.372 + 0631 \text{ CC.STRG.Gießleistung Rep. Meßwert})^2$ |
| 0.01131 | KalkVerbrauchSekMet |
| -712.4 | $(-2.372 + 0631 \text{ CC.STRG.Gießleistung Rep. Meßwert}) (-0.00002917 + \text{Einschluss_AL_AM})$ |

Abbildung 5: Darstellung eines quadratischen Modells. Erstellt mit dem machine learning framework for Mathematica.

In dieser Darstellung sind die einzelnen Faktoren entsprechend ihres potentiellen Einflusses auf die Vergießbarkeit gereiht; hier ein Produktinteraktionsterm zweier Variablen an erster Stelle, gefolgt von Einzeleinflüssen.

3.2 Anwendungsspezifische Herausforderungen

Eine besondere Herausforderung bei diesem Wissensextraktionsprozess ist die äußerst heterogene Beschaffenheit der Daten, die eine Wissensextrahierung basierend auf den Gesamtdaten aus verschiedenen Gründen unmöglich macht. Erstens ist aufgrund der teilweise unterschiedlichen Bauarten und Einstellungen der vier Stranggussanlagen ein Vergleich der auf den Stopfenverläufen basierenden Zielgröße zwischen den Anlagen unbrauchbar. Weiters ist es entscheidend, aus den vielen verschiedenen Werksmarken, Stahlmarken und Behandlungsrouten geeignete Szenarien zusammenzufassen, deren Analysen für die Experten neue Erkenntnisse liefern könnten. Es ist dafür eine enge Zusammenarbeit mit den Experten bei der Definition der Szenarien nötig. Auf diese Weise erhält man eine Vielzahl von Szenarien, die getrennt untersucht werden müssen. Zusätzlich ist es notwendig, die Datensätze für die etwa hundert verschiedenen Szenarien so aufzubereiten und zu bereinigen, dass interessante Einflussgrößen bei den Analysen zum Vorschein kommen und nicht von Artefakten überdeckt werden.

4 Automatisierung des Wissensextraktionsprozesses

Durch die vielen verschiedenen, zu untersuchenden Szenarien ist es notwendig, den Wissensextraktionsprozess effizient zu gestalten. Im Wesentlichen muss es möglich sein, für ein beliebiges Szenario die Schritte automatisch durchführen zu können, beginnend von der Datenaufbereitung über die Datenbereinigung hinweg zur Modellgenerierung und zu den entsprechenden Visualisierungen. Die erste Version einer entsprechenden Analyseumgebung ist bereits in der Forschungsabteilung der voest alpine Stahl implementiert, welche bereits automatisierte Analysen und eine flexible Definition von neuen Szenarien erlaubt. In Abbildung 6 ist beispielhaft eine Überblicksgrafik abgebildet, wie sie im Report eines automatischen Analysevorgangs zu finden ist.

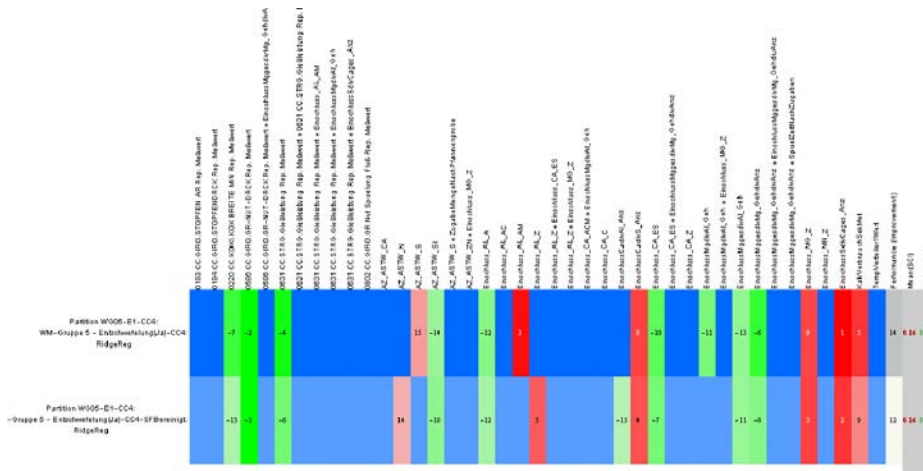


Abbildung 6: Überblicksgrafik zur Darstellung potentieller Einflussgrößen auf die Vergießbarkeit in einem bestimmten Szenario.

Diese Analyseumgebung basiert auf dem machine learning framework (mlf) für Mathematica, welches vom Software Competence Center Hagenberg in Kooperation mit der Fa. unisoftware plus gemeinsam entwickelt wird [Natschläger, Kossak, 2006]. Diese Software stellt die grundlegenden

Methoden des maschinellen Lernens zur Verfügung. Zur Zeit wird daran gearbeitet, die Analyseumgebung so zu gestalten, dass diese von den Stahl-Experten effizient für laufende betriebliche Analysen eingesetzt werden kann. Dazu gehört unter anderem, dass trotz des beträchtlichen Rechenaufwandes der Analysen die Ergebnisse möglichst rasch vorliegen. Aus diesem Grund ist die Analyseumgebung auf einem leistungsstarken Rechnercluster aufgesetzt, wodurch dank der Parallelisierungsfähigkeiten von Mathematica und mit einer Verteilung der Analysevorgänge auf mehrere Rechenknoten möglich ist.

5 Ergebnisse

Laut Aussagen der Stahl-Experten stellt die Analyseumgebung mit der gewählten Art der Darstellung der Modelle und der Datenvisualisierung ein übersichtliches und aussagekräftiges Werkzeug dar, wenn man sich in dessen prinzipiellen Darstellungsweisen einmal eingearbeitet hat. Die Analyseumgebung ist daher für prozessorientierte Forschungsingenieure ein wertvolles Hilfsmittel, Modellvorstellungen für das Cloggingverhalten zu entwickeln und potentielle Einflussgrößen für detaillierte betriebliche und / oder datenbasierte Untersuchungen zu orten.

Konkret ließen die umfangreichen, überblicksartigen Darstellungen der Datenanalysen für die verschiedenen Szenarien bereits generelle Einflussgrößen zutage treten. Unter diesen befindet sich eine direkt beeinflussbare Größe, nämlich die Gießgeschwindigkeit, deren Erhöhung immer einen positiven Einfluss auf die Vergießbarkeit zeigt. Das gefundene Verhalten konnte im Stahlwerk schon mit Erfolg praktisch nachgewiesen werden und wird bei Stahlsorten, bei denen keine anderen betrieblichen Beschränkungen dagegensprechen, bereits praktisch umgesetzt. Nachdem weitere Datenbestände in die Analysen integriert wurden, konnte ein weiterer genereller Einflussfaktor identifiziert werden: Eine höhere Anzahl von magnesiumhaltigen Einschlüssen im Stahl führt in fast allen Fällen zu einer Verschlechterung der Vergießbarkeit. Um dieses Erkenntnis zur Prozessverbesserung nutzen zu können, muss jedoch noch untersucht werden, ob es Betriebsparameter gibt, mit denen man die Anzahl der magnesiumhaltigen Einschlüsse in geeigneter Weise steuern kann.

6 Danksagung

Diese Arbeit wurde im Rahmen der COMET- und K-plus-Programme der österreichischen Forschungsförderungsgesellschaft unterstützt.

7 Literatur

- M. Drobnic and J. Himmelbauer. Creating comprehensible regression models. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, volume 11, number 5, pages 421-438, 2007.
- M. Drobnic. Inductive learning of fuzzy regression trees. *Proc. 4th Conf. of the European Society for Fuzzy Logic and Technology and 11 Recontres Francophones sur la Logique Floue et ses Applications (EUSFLAT/LFA 2005)*, pages 16-21, 2005.
- R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification (2nd ed.)*; John Wiley and Sons, 2001.
- T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*; Springer Verlag, 2001.
- J. Himmelbauer and M. Drobnic. Regularized numerical optimization of fuzzy rule bases. *Proc. 13th IEEE Int. Conf. on Fuzzy Systems*, pages 1655 - 1660, IEEE Press, 2004.

- B. Moser and U. Bodenhofer. Correspondences between fuzzy equivalence relations and kernels: Theoretical results and potential applications. *Proc. 15th IEEE Int. Conf. on Fuzzy Systems*, pages 10217-10223, IEEE Press. 2006.
- T. Natschläger, M. Drobics and F. Kossak. The machine learning framework for Mathematica: Creating interpretable computational models from data. *Proc. Wolfram Technology Conf.*: 2006.
- T. Natschläger. *Iterative polynomial regression for high dimensional data*. Technical Report, Software Competence Center Hagenberg, 2008. In preparation.
- T. Natschläger, F. Kossak and M. Drobics. Extracting knowledge and computable models from data - Needs, expectations, and experience. *Proc. 13th IEEE Int. Conf. on Fuzzy Systems*, pages 493 - 498, IEEE Press. 2004.
- L. Trueba, K. D. Peaslee, J. D. Smith. Simulation of Nozzle Clogging During the Continuous Casting of Aluminum-killed Steel; *AIST Transactions*, page 123, 2006.