Bestimmung von Isothermenparametern mit Hilfe des maschinellen Lernens

Sven Felske[1], Oliver Ritthoff[2] und Ralf Klinkenberg[2]

[1] Universität Dortmund, Fachbereich Chemietechnik, Lehrstuhl für Anlagentechnik, 44221 Dortmund
E-mail: sven.felske@ct.uni-dortmund.de
[2] Universität Dortmund, Fachbereich Informatik, Lehrstuhl für künstliche Intelligenz, 44221 Dortmund
E-mail: {ritthoff, klinkenberg}@ls8.cs.uni-dortmund.de

Zusammenfassung Die Modellierung eines chromatographischen Prozesses ermöglicht die Bestimmung der optimalen Betriebsparameter für die Trennung der Komponenten eines Stoffgemisches mittels modellbasierter Optimierung. Voraussetzung für eine genaue Modellierung ist dabei die Kenntnis der stoffabhängigen Parameter des zugrunde liegenden physikalischen Modells, insbesondere der Adsorptionsisotherme. Die messtechnische Bestimmung der Isotherme ist zeit- und materialaufwändig und deshalb nicht generell durchführbar. Eine Bestimmung der Isothermenparameter aus Chromatogrammen mittels mathematischer Parameterschätzung hingegen senkt diesen Aufwand, hat jedoch den Nachteil der Abhängigkeit von guten Startwerten. Die hier vorgestellte Methode nutzt die Approximationsfähigkeiten von Support Vector Machines zur Bestimmung von Isothermenparametern aus wenigen Merkmalen von Chromatogrammen und erfordert derartige Startwerte nicht.

1 Grundlagen der Chromatographie

Die Chromatographie ist ein thermisches Trennverfahren zur Separation mehrerer Komponenten eines Stoffgemisches in einem fluiden Fördermittel, wobei das Fluid flüssig (Flüssigchromatographie) oder gasförmig (Gaschromatographie) sein kann. Die Chromatographie wird vor allem bei schwierigen Trennungen mit Produkten hoher Wertschöpfung, wie sie in der Pharmazie und Biotechnologie anfallen, angewendet. Grundlage dieser Trennung ist die selektive Adsorption bzw. Desorption einer Komponente im Fluid an einen Feststoff. Adsorption ist die Anlagerung von Teilchen (Adsorptiv) an die Oberfläche einer festen Phase (Adsorbens). Die Anlagerung erfolgt durch molekulare Wechselwirkungen zwischen Adsorbtiv und Adsorbens. Die Entfernung der adsorbierten Teilchen (Adsorpt) vom Adsorbens wird Desorption genannt. Die Desorption erfolgt durch die Verdrängung mittels eines stärker adsorbierbaren Teilchens, durch Regenerierung des Adsorbens mit einem inerten Fluid (Eluent) oder durch Energiezufuhr [Rut84] [Gui94]. Es stellt sich zwischen der Konzentration des Adsorptivs und der Beladung am Adsorbens ein thermodynamisches Gleichgewicht ein, welches durch die Adsorptionsisotherme beschrieben wird. Diese ist abhängig von den Konzentrationen aller adsorbierbaren Komponenten, vom Adsorbens und dem verwendeten Eluenten. Die diskontinuierliche Trennung eines binären Gemisches in einer Chromatographiesäule wird in Abbildung 1 dargestellt.

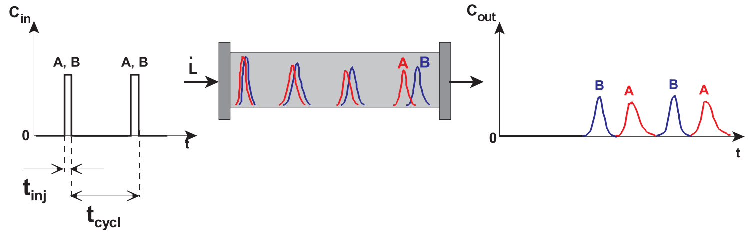

Abbildung 1. Prinzip der Batch-Chromatographie

In der Trennsäule befindet sich eine Packung aus porösen, sehr kleinen Feststoffpartikeln (Adsorbens), die von einem inerten Fluid durchströmt werden. Die Stofftrennung ergibt sich aus den unterschiedlichen Verweilzeiten der einzelnen Komponenten im Adsorbens. Aus den unterschiedlichen Affinitäten der Komponenten zum Adsorbens resultieren verschiedene Wanderungsgeschwindigkeiten der Konzentrationsfronten der Komponenten des Fluids. Der Aufgabeimpuls wird beim Durchwandern der Packung in zwei Peaks aufgeteilt. Der Peak der stärker adsorbierbaren Komponente (A) folgt durch die langsamere Wanderungsgeschwindigkeit dem Peak der schwächer adsorbierbaren Komponente (B). Nach der Zykluszeit wird erneut ein Gemisch aus (A) und (B) in die Säule injiziert.

2 Modellierung chromatographischer Prozesse

Die Modellierung chromatographischer Prozesse ermöglicht deren Optimierung, abhängig von der Güte des verwendeten Modells und dessen Parametern. Mit der Anzahl der im Modell berücksichtigten Effekte steigt die Genauigkeit der Simulation, aber auch die Anzahl der zu bestimmenden Modellparameter.

Für eine möglichst exakte Simulation wird das *General Rate Model* verwendet [Gu95] . Es berücksichtigt neben Konvektion und Dispersion die stoffabhängigen Einflüsse des Stofftransportwiderstands zwischen Fluid und Partikel, der Diffusion der Komponenten im Partikel und der Gleichgewichtsadsorption.

Aus dieser Bilanzierung ergibt sich als Differentialgleichung

$$\frac{\partial c_i}{\partial t} = D_{ax}\frac{\partial^2 c_i}{\partial x^2} - u\frac{\partial c_i}{\partial x} - \frac{3\,(1-\varepsilon)\,k_{Film,i}}{\varepsilon r_p}\,(c_i - c_{pi}(r_p)) \tag{1}$$

für die fluide Phase und

$$(1-\varepsilon_p)\,\frac{\partial q_i}{\partial t} + \varepsilon_p\frac{\partial c_{pi}}{\partial t} = \varepsilon_p D_{p,i}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c_{pi}}{\partial r}\right)\right] \tag{2}$$

für die Partikel[3]. Die Gleichung für die Adsorptionsisotherme ist nicht vorgegeben und muss festgelegt werden. Je nach Adsorptionsverhalten gibt es dabei verschiedene Gleichungsansätze. Eine gängige Beziehung für Stoffsysteme mit Adsorptionsmaximum ist die Langmuir-Isotherme in der allgemeinen Form

$$q_i = \frac{H_i \, c_{pi}}{1 + \sum\limits_j L_{i,j} \, c_{pj}} \quad . \tag{3}$$

3 Die Stützvektormethode

Die *Stützvektormethode (engl. Support Vector Machine, SVM)* basiert auf dem Prinzip der strukturellen Risikominimierung (SRM) aus der statistischen Lerntheorie [Vap98]. Die Idee der strukturellen Risikominimierung ist es, die Hypothese h aus dem Hypothesenraum H zu finden, für die der kleinste echte Fehler probabilistisch garantiert werden kann. Der echte Fehler einer Hypothese h ist bei Klassifikationsproblemen die Wahrscheinlichkeit, bei der Klassifikation eines ungesehenen, zufällig gezogenen Testbeispiels einen Fehler zu machen. Die folgende Schranke verbindet den echten Fehler einer Hypothese h mit ihrem Fehler auf den Trainingsbeispielen und der Komplexität von h [Vap98]. Sei δ die maximale Wahrscheinlichkeit, dass diese Schranke falsch ist, also der Fehler höher als angegeben ist.

$$P(error(h)) \leq train\_error(h) + 2\sqrt{\frac{d(ln\frac{2n}{d} + 1) - ln\frac{\delta}{4}}{n}} \tag{4}$$

Die Schranke gilt mit einer Wahrscheinlichkeit von $1-\delta$. n ist die Anzahl der Trainingsbeispiele und d ist die VC-Dimension (VCdim) [Vap98]. Die VC-Dimension ist ein Maß für die Komplexität des Hypothesenraums, aus dem h gezogen wurde. Die obige Fehlerschranke reflektiert den wohlbekannten Trade-off zwischen der Komplexität des Hypothesenraums und dem Trainingsfehler. Ein kleiner Hypothesenraum (kleine VCdim) wird wahrscheinlich keine gute Hypothese enthalten und so zu einem hohen Trainingsfehler (und echten Fehler) führen. Andererseits wird ein zu großer Hypothesenraum (hohe VCdim) wahrscheinlich eine Hypothese mit kleinem Trainingsfehler enthalten, aber der zweite Summand auf der rechten Seite der Ungleichung wird groß. Diese Situation wird allgemein als „Overfitting“ bezeichnet - niedriger Trainingsfehler, aber schlechte Generalisierung auf neuen Beispielen. Die Wahl eines Hypothesenraums geeigneter Komplexität ist somit entscheidend.

Das SRM-Prinzip berücksichtigt die Komplexität des Hypothesenraums, indem es eine Struktur über den Hypothesenraum legt. Diese Struktur gliedert den Hypothesenraum in Teilmengen steigender VC-Dimension. Sei d_i die VCdim des Hypothesenraums H_i:

$$H_1 \subset H_2 \subset H_3 \subset ... \subset H_i \subset ... \qquad \text{und} \qquad \forall i : d_i \leq d_{i+1} \tag{5}$$

[3] Die Erläuterung der Variablen ist im Symbolverzeichnis am Ende des Artikels zusammengefasst.

Es werden der Hypothesenraum H_i und die Hypothese h gewählt, die die obige Fehlerschranke minimieren. Bemerkenswert ist, dass die Dimension des Attributraums nicht in diesem Zusammenhang auftaucht. Somit kann die SVM potentiell auch in sehr hochdimensionalen Räumen gut generalisieren.

Die Stützvektormethode ist ein sehr universelles Lernverfahren zur Klassifikation, Regression und Dichteschätzung. Durch den Einsatz von Mercer-Kernels können SVMs so parametrisiert werden, dass neben linearen Approximatoren auch Polynomklassifikatoren, Radial Basis Netze und zweilagige neuronale Feed-Forward-Netze realisierbar sind [BGV92]. Neben den bemerkenswerten lerntheoretischen Eigenschaften ist die SVM auch aus komplexitätstheoretischer Sicht sehr interessant. Das Training einer SVM lässt sich auf ein semi-definites quadratisches Optimierungsproblem zurückführen, das sowohl theoretisch als auch praktisch effizient global optimal gelöst werden kann.

4 Parameterbestimmung mittels maschinellem Lernen

Die Idee hierbei ist die Bestimmung der Isothermenparameter aus der Position und Form der Konzentrationspeaks. Im Gegensatz zu einem mathematischen Parameterschätzer, welcher ausgehend von den Startwerten die Anpassung der simulierten Profile an den vorgegebenen Profilen sucht, wird hier je eine SVM pro Parameter mit simulierten Peaks trainiert und auch getestet, wobei ein RBF-Kernel der Form

$$y = \frac{1}{e^{\gamma \|x - x_0\|^2}} \tag{6}$$

verwendet wird.[4] Die SVM approximiert dabei die Abhängigkeit des zu ermittelnden Parameters von der Position und Form des Konzentrationspeaks. Voraussetzung für eine gute Bestimmbarkeit der Parameter ist, dass diese auch einen Einfluss auf das Konzentrationsprofil haben. Somit werden hier symmetrische Multi-Langmuir Gleichungen der Form

$$q_i = \frac{H_i \, c_{pi}}{1 + \sum\limits_j L_j \, c_{pj}} \quad , \tag{7}$$

verwendet. Eine weitere Voraussetzung ist die Wahl einer repräsentativen Trainingsmenge in den Grenzen des Anwendungsbereichs, in diesem Fall eine Normalverteilung der Beispiele. Die verwendeten Trainingsgrenzen sind in Tabelle 1 dargestellt. Die Grenzen der Beispiele sind hier so gewählt, dass ein einfaches nichtlineares Trennproblem vorliegt und die erste Komponente immer schwächer adsorbiert, als die zweite.

Analog zu neuronalen Netzen kann die trainierte SVM im Idealfall innerhalb der gewählten Trainingsgrenzen die zu einem Profil gehörenden Parameter bestimmen.

[4] Eine entsprechende Erläuterung der verwendeten SVM-Parameter findet sich im Symbolverzeichnis am Ende dieses Artikels.

Tabelle 1. Trainingsgrenzen der stoffabhängigen Parameter

	untere Grenze	*obere Grenze*
H_1	5	10
H_2	18	23
L_1	50	250
L_2	250	450
k_{Film}	1e-4	5e-4

Eine Datenreihe für die Konzentrationsprofile besteht aus mehreren hundert oder tausend Datenpunkten, bei denen zudem viele eine Konzentration von Null haben. Die Verwendung der gesamten Konzentrationsreihe als Trainingsbeispiel lieferte große Approximationsfehler [DDE$^+$02]. Ursache hierfür ist die fehlende Robustheit dieser Daten, da die aus den variierenden Parametern resultierenden Unterschiede sich teilweise nur in einigen kleinen Bereichen der Datenreihe zeigen. Daher wurde getestet, ob eine Reduzierung der Dimensionalität der Beispiele auf wenige markante Punkte ausreicht, um höhere Genauigkeiten zu erzielen. In Anlehnung an die Funktionsanalyse werden die Wendepunkte und das Maximum der Peaks als Eingangsgrößen verwendet, welche die notwendige Robustheit besitzen.

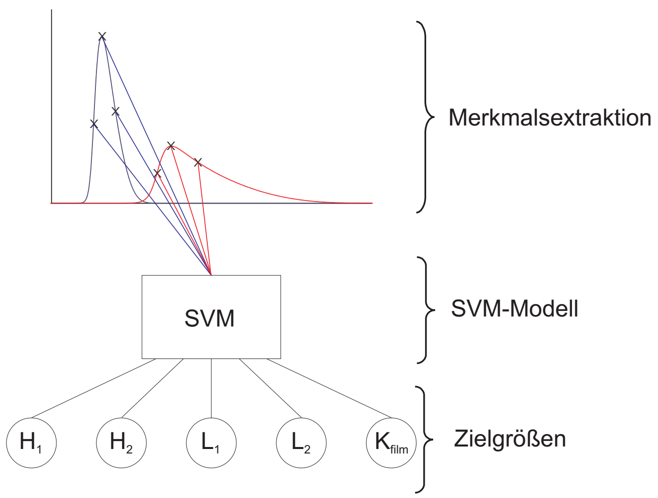

Abbildung 2. Merkmalsgenerierung aus Konzentrationsprofilen

Abbildung 2 zeigt das Prinzip der Parameterbestimmung. Aus einem Chromatogramm werden die Merkmale als Eingangsgrößen extrahiert. Die Zielgröße der SVM ist jeweils ein Isothermenparameter oder der Stofftransportwiderstand. Die entsprechenden Experimente wurden mit der flexiblen Lernumgebung YALE[5] durchgeführt [RKF+01] [FKMR02]. Skriptbasiert werden hierbei die Test- und Trainingsdaten der Parameter in ein Simulationsprogramm eingelesen und aus den daraus resultierenden Konzentrationsprofilen die Merkmale bestimmt. Mit den Trainingsdaten werden die Support Vector Machines trainiert und mit den entsprechenden Merkmalen der Testdaten die Genauigkeit überprüft.

5 Ergebnisse

5.1 Bestimmung der SVM-Parameter

Zur Abschätzung der SVM-Parameter e, γ und C wurde mit den in Tabelle 2 genannten Werten eine Parameterstudie für die Isothermenparameter der ersten Komponente und k_{Film} durchgeführt.

Tabelle 2. Parameterstudie zur Abschätzung der optimalen SVM-Parameter

C	1	10	100	1000
γ	0,01	0,1	1	2
e	0,001	0,01	0,1	1

Insgesamt ergeben sich 64 getestete SVM-Einstellungen pro zu lernender Zielgröße. Training und Test erfolgten über 4-fache Kreuzvalidierung auf 1200 Beispielen. Aus dieser Studie ergibt sich, dass die SVM-Parameter einen deutlichen Einfluss auf die resultierende Güte des Lernens haben. Der durchschnittliche relative Fehler schwankt bei den Parameterkombinationen um eine Größenordnung. Die Parameterkombinationen mit dem geringsten Fehler sind in Tabelle 3 dargestellt.

Bei der Bestimmung von k_{Film} traten aufgrund der kleinen Dimension des Parameters sehr große relative Fehler auf. Als Ergebnis wurde generell der ungefähre Mittelwert der Trainingsgrenzen ausgegeben, unabhängig von den SVM-Parametern. Da die Dimension von k_{Film} mit Werten zwischen $1e-4$ und $5e-4$ deutlich kleiner ist, als bei den anderen Parametern, wurden die Werte skaliert, indem k_{Film} mit $1e4$ multipliziert wurde. Die daraus resultierenden Fehler sind in Tabelle 3 dargestellt.

Die Ergebnisse der Parameterstudie für die Isothermenparameter der ersten Komponente sind in den Abbildungen 3 und 4 zusammengefasst. Die Achsen

[5] http://yale.cs.uni-dortmund.de

Tabelle 3. Die optimalen SVM-Parameter der Parameterstudie

Zielgröße	γ	e	C	*relativer Fehler* [%]
H_1	0,01	0,01	1000	0,281
H_2	0,01	0,01	100	0,120
L_1	0,01	1	1000	1,438
L_2	0,01	1	1000	0,519
k_{Film}	0,01	0,01	100	2,365

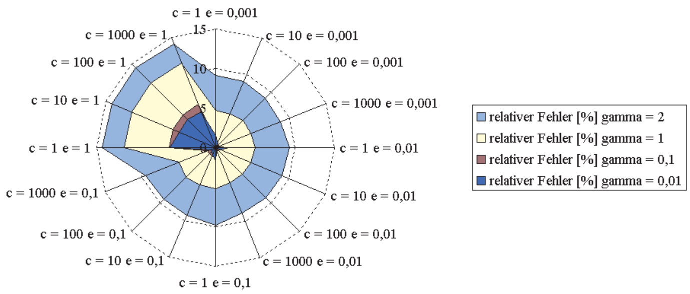

Abbildung 3. Einfluss der SVM-Parameter auf die Bestimmung von H_1

nennen jeweils C und ϵ und die Fläche zeigt γ einer getesteten SVM-Parameter-Kombination. Aufgetragen ist der mittlere relative Fehler der Zielgrößen bei diesen SVM-Parametern. Der Einfluss der SVM-Parameter ist bei den Zielgrößen unterschiedlich. Die Erhöhung der absoluten Fehlertoleranz e führt außer bei L ab einem Wert von 1 zu einem deutlichen Anstieg des Fehlers, da der zulässige Fehler nun die gleiche Dimension wie die zu bestimmenden Werte hat. Große Werte für γ führen zu sehr schmalen Glockenkurven, mit denen eine gute Approximation durch die steilen Fronten der Kurve schwierig wird. Bei den Isothermenparametern H und L sowie beim Stofftransportwiderstand k_{Film} wird mit kleinen Werten für γ die beste Genauigkeit erreicht. Erst bei einer guten Kombination der anderen SVM-Parameter wird der Einfluss der Komplexität C des gelernten Modells deutlich. Es werden generell komplexere Modelle für eine gute Approximation benötigt.

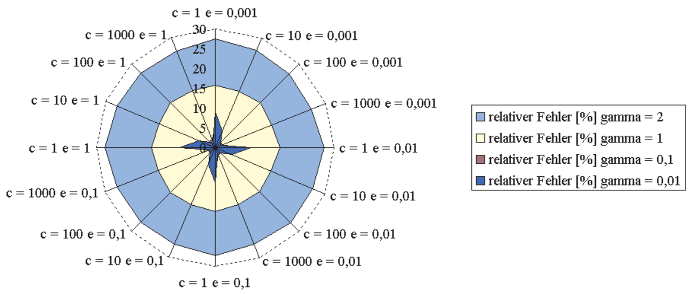

Abbildung 4. Einfluss der SVM-Parameter auf die Bestimmung von L_1

5.2 Einfluss der Trainingsdatenanzahl

Entscheidend für ein erfolgreiches Lernen der SVM ist eine ausreichende Anzahl an Lernbeispielen, in diesem Fall von simulierten Konzentrationsprofilen. Mit Zunahme der Beispiele steigt die Genauigkeit, aber auch die benötigte Trainings- und Simulationszeit. In der Abbildung 5 sind die aus 800 Testdaten ermittelten mittleren relativen Fehler für verschiedene Trainingsmengen dargestellt.

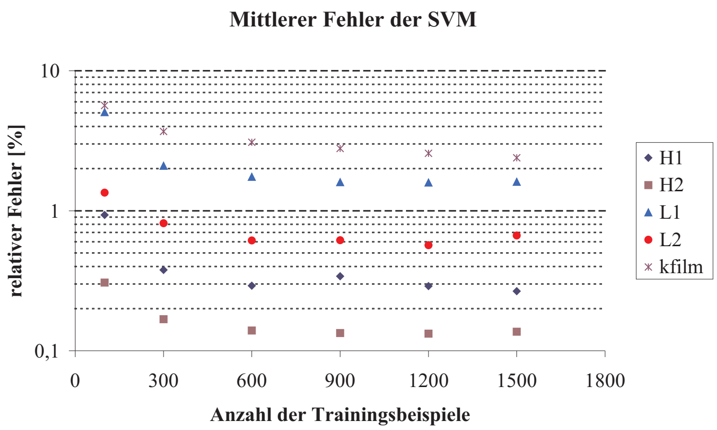

Abbildung 5. Einfluss der Anzahl der Trainingsbeispiele auf den relativen Fehler

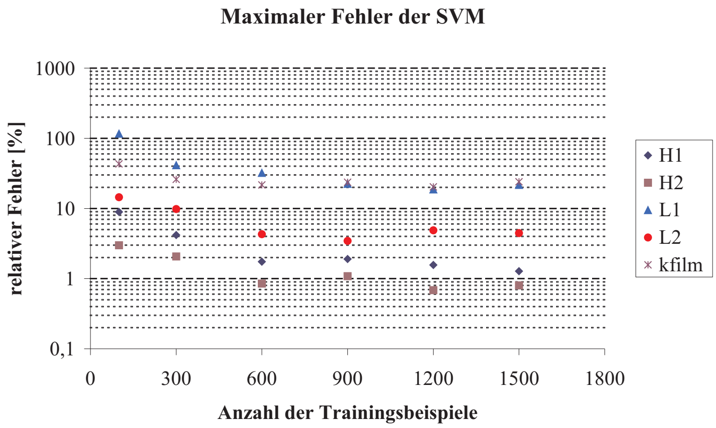

Abbildung 6. Abhängigkeit des maximalen Fehlers von der Trainingsdatenmenge

Generell zeigt sich, dass die Parameter des zweiten Peaks besser bestimmt werden können. Durch die längere Verweilzeit in der Säule können sich die aus den Isothermenparametern ergebenen Unterschiede deutlicher ausprägen, als bei dem ersten Peak. Auch die Lage der Trainingsgrenzen begünstigt ein besseres Lernen, da die aus der Variation der Parameter resultierenden Veränderungen des Peaks deutlicher ausgeprägt sind, als bei den Trainingsgrenzen für den ersten Peak. Bei den Parametern selbst verhält sich die Größe des relativen Fehlers entsprechend zu dem Einfluss der Parameter auf die resultierenden Profile. Den größten Einfluss haben die Henrywerte H und sind somit auch am besten zu lernen. Den geringsten Einfluss hat der Stofftransportwiderstand k_{film}, welcher auch den größten Fehler hat. Eine Erhöhung der Trainingsmenge von 300 auf 600 Beispielen führt insbesondere bei k_{Film} und dem Langmuirkoeffizienten L_1 zu einer Verbesserung der Genauigkeit. Eine weitere Erhöhung der Beispielmenge auf bis zu 1500 führt nur noch bei k_{film} zu einer Reduzierung des mittleren Fehlers. Ansonsten ändert sich der mittlere Fehler nur noch gering. Mit Ausnahme von k_{film} liefern zusätzliche Trainingsdaten keinen weiteren Informationsgewinn für das Training der SVM, ab 600 Trainingsbeispielen erreicht die SVM für diese Parameter ihre mögliche Genauigkeit.

Abbildung 6 zeigt die maximalen Fehler aus den 800 Testdaten. Diese liegen generell um ein vielfaches höher, als die mittleren Fehler. Im Trainingsraum gibt es Regionen, welche von der SVM deutlich schlechter beherrscht werden. Bei der Erhöhung von 300 auf 600 Daten werden die maximalen Fehler noch deutlich reduziert, ab 900 Trainingsdaten ist auch hier keine einheitliche Tendenz mehr vorhanden.

Die mittleren Fehler liegen hier in vertretbaren Bereichen, da die Abweichungen der resultierenden Profile von der Vorlage gering sind. Abbildung 7 zeigt an einem Beispiel die Übereinstimmung der Simulation mit den von der SVM be-

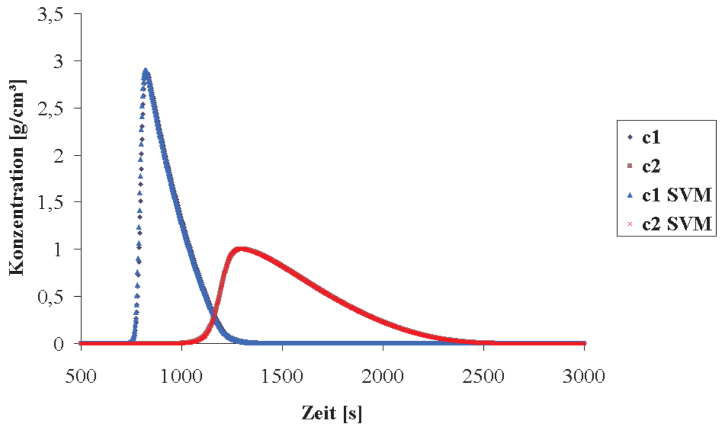

Abbildung 7. Vergleich der Vorgabe mit den aus den Parametern der SVM resultierenden Profilen

stimmten (c1 SVM, c2 SVM) und den tatsächlichen Parametern (c1, c2). Die Abweichungen der Parameter haben nur einen geringen Einfluss auf das Konzentrationsprofil, wodurch die Kurven fast deckungsgleich sind.

6 Zusammenfassung und Ausblick

Mit der hier vorgesellten Methode können aus wenigen Datenpunkten die stoffabhängigen Parameter für eine Langmuir-Isotherme bestimmt werden. Bei den hier gewählten Trainingsgrenzen der Parameter einer Langmuir-Isotherme reichen für eine gute Bestimmung 600 Trainingsbeispiele. Bei einigen Beispielen können aber deutliche Fehler auftreten. Da die SVM nur für die stoffabhängigen Parameter trainiert wird, dürfen sich die sonstigen Anlagen- und Betriebsparameter nicht ändern. Die daraus resultierenden Veränderungen können sonst nicht richtig interpretiert werden. Eine Erweiterung der zu lernenden Parameter bietet so ein größeres Einsatzspektrum, wobei der Trainingsaufwand durch die zusätzliche Dimension ansteigen wird. Ebenfalls zu Fehlern kommt es, wenn die gesuchten Parameter außerhalb der Trainingsgrenzen liegen. Der Einfluss der Trainingsgrenzen auf die benötigte Trainingsdatenmenge und die resultierenden Fehler ist noch zu untersuchen.

Ein Schwachpunkt bezüglich der Anwendung auf gemessene Profile ist die Verwendung der Wendepunkte als Merkmal. Wenn die Messdaten stark rauschen ist deren exakte Bestimmung sehr fehleranfällig. Dies kann als Folge zu falschen Ergebnissen führen.

Der Vorteil gegenüber einem mathematischen Parameterschätzer ist die Unabhängigkeit von Startwerten. Eine Kombination aus SVM für die Generierung von Startwerten und einem mathematischen Parameterschätzer könnte somit zu sehr guten Resultaten führen. Dies wird in folgenden Arbeiten untersucht werden.

7 Danksagung

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Sonderforschungsbereichs Computational Intelligence (SFB 531) an der Universität Dortmund gefördert.

8 Symbolverzeichnis

	Symbol
c_i	Konzentration der Komponente i in der fluiden Phase
c_{Pi}	Konzentration der Komponente i im Partikel
D_{ax}	Dispersionskoeffizient
D_{Pi}	Diffusionskoeffizient der Komponente i im Partikel
H_i	Henrykoeffizient der Komponente i
$k_{film,i}$	Stoffübergangskoeffizient der Komponente i
L_i	Langmuirkoeffizient der Komponente i
q_i	Gleichgewichtsbeladung der Komponente i auf der Partikeloberfläche
r_P	Partikelradius des Packungsmaterials
u	Zwischenkorngeschwindikeit des Fluids
ϵ	Porosität der Packung
ϵ_P	Porosität der Partikel
e	Grad der erlaubten Abweichung des vorhergesagten
	vom echten Funktionswert
C	obere Schranke für den maximalen Einfluss eines einzelnen Trainingsbeispiels
γ	Breite der Gaußkurve eines gegebenen RBF-Kernels

Literatur

[BGV92] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A traininig algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT-92)*, pages 144–152, 1992.

[DDE+02] G. Daniel, J. Dienstuhl, S. Engell, S. Felske, K. Goser, R. Klinkenberg, K. Morik, O. Ritthoff, and H. Schmidt-Traub. Novel learning tasks, optimization, and their application. In H.-P. Schwefel, I. Wegener, and K. Weinert, editors, *Advances in Computational Intelligence – Theory and Practice*, Natural Computing Series, chapter 8, pages 245–318. Springer, Berlin, 2002.

[FKMR02] S. Fischer, R. Klinkenberg, I. Mierswa, and O. Ritthoff. YALE: Yet Another Learning Environment - Tutorial. Technical Report CI-136/02, Sonderforschungsbereich 531, Universität Dortmund, Juni 2002. *http://yale.cs.uni-dortmund.de*.

[Gu95] T. Gu. *Mathematical Modelling and Scale Up of Liquid Chromatography.* Springer, New York, USA, 1995.

[Gui94] G. Guiochon. *Fundamentals of Preparative and Nonlinear Chromatography.* Academic Press, Boston, USA, 1994.

[RKF+01] O. Ritthoff, R. Klinkenberg, S. Fischer, I. Mierswa, and S. Felske. YALE: Yet Another Machine Learning Environment. In R. Klinkenberg, S. Rüping, A. Fick, N. Henze, C. Herzog, R. Molitor, and O. Schröder, editors, *LLWA 01 – Tagungsband der GI-Workshop-Woche* Lernen – Lehren – Wissen – Adaptivität, pages 84–92. Technischer Bericht Nr. 763, Fachbereich Informatik, Universität Dortmund, Oktober 2001. *http://yale.cs.uni-dortmund.de*.

[Rut84] D. M. Ruthven. *Principles of Adsorption and Adsorption Processes.* John Wiley and Sons Inc., New York, USA, 1984.

[Vap98] Vladimir N. Vapnik. *Statistical Learning Theory.* Wiley, Chichester, GB, 1998.