

Diplomarbeit

Simulation einer Multikapillarsäule

Elisabeth Böhmer

23. Februar 2015

Betreuer:

Prof. Dr. Sven Rahmann

Prof. Dr. Jörg Rahnenführer

Fakultät für Informatik

Algorithm Engineering (Ls11)

Technische Universität Dortmund

<http://ls11-www.cs.tu-dortmund.de>

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation und Hintergrund	1
1.2	Aufbau der Arbeit	1
2	Grundlagen	3
2.1	Chromatographie	3
2.2	Probabilistische Arithmetische Automaten	3
2.2.1	Definition eines PAA	4
2.2.2	Algorithmus zur Berechnung der Wartezeit	4
3	Modell	5
3.1	2-Parameter Modell	5
3.1.1	PAA zu diesem Modell	5
3.1.2	Probleme mit dem Modell	5
3.2	Weiteres Modell	5
4	Implementierung	7
4.1	Simulation	7
4.1.1	Step-by-step	7
4.1.2	By-Event	7
4.2	PAA	7
5	Evaluation	9
6	Diskussion	11
A	Weitere Informationen	13
	Abbildungsverzeichnis	15
	Algorithmenverzeichnis	17
	Literaturverzeichnis	19

Kapitel 1

Einleitung

1.1 Motivation und Hintergrund

1.2 Aufbau der Arbeit

Im Rahmen der Diplomarbeit soll eine Multikapillarsäule simuliert werden.

Es geht dabei nicht um eine physikalische Simulation auf molekularer Ebene, sondern um die Entwicklung eines abstrakten, probabilistischen Modells, welches mit möglichst wenig Parametern auskommt.

In der Simulation soll jeweils nur ein Stoff, charakterisiert durch diese Parameter, simuliert werden. Das Resultat einer Simulation ist jeweils ein einzelner Peak, welcher durch seine Lage und Form beschrieben werden kann. Die Lage entspricht dabei der Retentionszeit am Maximum des Peaks. Die Form ist durch seine Breite an einer bestimmten Höhe (zum Beispiel Halbwertsbreite) sowie Schiefe gekennzeichnet.

Umgekehrt soll es möglich sein, für eine gegebene Peaklage und Peakform die nötigen Simulationsparameter zu ermitteln, mit denen ein solcher Peak simuliert werden kann.

Zum Vergleich liegen einige MCC-IMS-Messungen von Mischungen sieben bekannter Stoffe vor. Diese Datensätze wurden zur Verfügung gestellt von der Firma B & S Analytik (<http://www.bs-analytik.de/>). Zu den dadurch gegebenen Peaks sollen Parameter für die Simulation ermittelt und diese Peaks simuliert werden.

Gesucht ist letztendlich eine allgemeine Entsprechung der Simulationsparameter zu den Parametern mit denen ein Peak beschrieben werden kann, falls diese existiert. Eine Berechnung des Peaks nur durch die Simulationsparameter und umgekehrt eine Vorhersage der Simulationsparameter zu bekannten Peakdaten soll damit realisiert werden.

Kapitel 2

Grundlagen

2.1 Chromatographie

Die Gaschromatographie (GC) ist ein Verfahren, mit dem gasförmig vorliegende Stoffgemische aufgetrennt oder analysiert werden können. Die Auftrennung erfolgt dabei zwischen zwei sogenannten Phasen, der stationären und der mobilen Phase, welche sich in unterschiedlichen Aggregatzuständen befinden und untereinander nicht mischen.

Beispielsweise kann die GC in einer Multikapillarsäule (MCC, engl. Multi Capillary Colum) stattfinden. Sie besteht aus ca. 1000 einzelnen Kapillaren. Jede davon ist innen mit der sog. stationären Phase beschichtet. Außerdem kommt ein Trägergas, die sog. mobile Phase, zum Einsatz, welches die Analyte durch die Säule transportiert. Die Substanzen unterscheiden sich vor allem durch ihre Wechselwirkungen mit der stationären Phase. Während dieser Wechselwirkungen haften die Teilchen an der stationären Phase, bewegen sich also nicht fort. Finden wenig Wechselwirkungen statt, passieren die Teilchen die Säule schneller, als wenn viele Wechselwirkungen stattfinden. Dies beeinflusst die Retentionszeit, also die Zeit, die zum Durchlaufen der Säule gebraucht wird.

Nach Durchlaufen der Säule können die Substanzen mit anderen Verfahren wie Massenspektrometrie (MS) oder Ionen-Mobilitäts-Spektrometrie (IMS) weiter analysiert werden.

Zu beobachten ist, dass schnelle Teilchen Peaks zu frühen Zeitpunkten erzeugen, die eine relativ geringe Varianz aufweisen, hingegen spätere Peaks tendenziell breiter werden. Ideale Peaks haben die Form einer Gaußkurve, jedoch tritt oft ein Tailing auf. Dieses wird unter anderem verursacht durch Adsorptionseffekte, die beim Altern einer Säule auftreten [1].

2.2 Probabilistische Arithmetische Automaten

Ein Probabilistischer Arithmetischer Automat (PAA) nach [2] ist ein Modell, mit dem eine Folge zufälliger Operationen beschrieben werden kann. Für PAA existieren Algorithmen,

welche eine gemeinsame Verteilung von Zuständen und Werten oder auch die Verteilung der Wartezeit für einen Wert berechnen. Wie in Kapitel 3 beschrieben wird, kann das Modell zur Simulation einer Multikapillarsäule auch als PAA formuliert werden. Mit dieser Formulierung ist die Zeit, die zum Durchlaufen einer Säule gebraucht wird, dann die Wartezeit für den Wert, welcher der Länge der Säule entspricht. Deshalb kann ein PAA nützlich sein, um neben der eigentlichen Simulation auch noch eine erwartete Verteilung der Ankunftszeiten der Teilchen zu berechnen.

Zunächst sei hier eine Definition für den PAA gegeben, anschließend wird der Algorithmus zur Berechnung der Wartezeit beschrieben.

2.2.1 Definition eines PAA

2.2.1 Definition (PAA). Ein Probabilistischer Arithmetischer Automat (PAA) ist ein Tupel $\mathcal{P} = (\mathcal{Q}, q_0, T, \mathcal{V}, v_0, \mathcal{E}, (e_q)_{q \in \mathcal{Q}}, (\theta_q)_{q \in \mathcal{Q}})$, dabei ist:

- \mathcal{Q} eine endliche Menge von Zuständen
- $q_0 \in \mathcal{Q}$ der Startzustand
- $T : \mathcal{Q} \times \mathcal{Q} \rightarrow [0, 1]$ eine Übergangsfunktion mit $\sum_{q' \in \mathcal{Q}} T(q, q') = 1$ das heißt $(T(q, q'))_{q, q' \in \mathcal{Q}}$ ist eine stochastische Matrix
- \mathcal{V} eine Menge von Werten
- v_0 der Startwert
- \mathcal{E} eine endliche Menge von Emissionen
- $e_q : \mathcal{E} \rightarrow [0, 1]$ eine Wahrscheinlichkeitsverteilung der Emissionen für jeden Zustand
- $\theta_q : \mathcal{V} \times \mathcal{E} \rightarrow \mathcal{V}$ eine Operation für jeden Zustand

Dabei entspricht $\mathcal{P} = (\mathcal{Q}, q_0, T)$ einer Markovkette und $\mathcal{P} = (\mathcal{Q}, q_0, T, \mathcal{E}, (e_q)_{q \in \mathcal{Q}})$ einem Hidden Markov Model.

2.2.2 Algorithmus zur Berechnung der Wartezeit

Kapitel 3

Modell

3.1 2-Parameter Modell

Als Grundlage für die Entwicklung eines Simulationsmodells dient die Beobachtung des chromatographischen Prozesses. Dieser ist wie in TODO beschrieben gekennzeichnet durch viele Teilchen, welche häufig zwischen der stationären und mobilen Phase wechseln. Es liegt also nahe, dass im Modell ebenfalls Teilchen simuliert werden, die zwischen zwei Zuständen, welche die beiden Phasen repräsentieren, wechseln. Der Phasenwechsel geschieht jeweils mit einer bestimmten Wahrscheinlichkeit. Dabei kann es möglich sein, dass die Wahrscheinlichkeiten für den Wechsel in die eine oder andere Richtung voneinander abhängig sind, jedoch muss dies nicht sein, sodass für das Modell der allgemeinere Fall von unabhängigen Wahrscheinlichkeiten angenommen wird. Es sei also p_s die Wahrscheinlichkeit, dass ein Teilchen, welches sich bereits in der stationären Phase befindet, auch stationär bleibt und $1 - p_s$ die Wahrscheinlichkeit, dass es in die mobile Phase übergeht. Analog seien p_m und $1 - p_m$ die Wahrscheinlichkeiten, dass ein mobiles Teilchen in der mobilen Phase bleibt bzw. zur stationären Phase wechselt.

Damit ergibt sich als Modell ein einfacher Automat mit zwei Zuständen $\mathcal{Q} = s, m$

Für die Simulation müssen x Teilchen den Automaten durchlaufen, dabei Ort festhalten für abbruchbedingung und zeit für peak Das ganze 1000 mal wiederholen da je nur eine kapillare simuliert wurde

3.1.1 PAA zu diesem Modell

Text

Und noch mehr Text, der sich auf 3.1 bezieht

3.1.2 Probleme mit dem Modell

3.2 Weiteres Modell

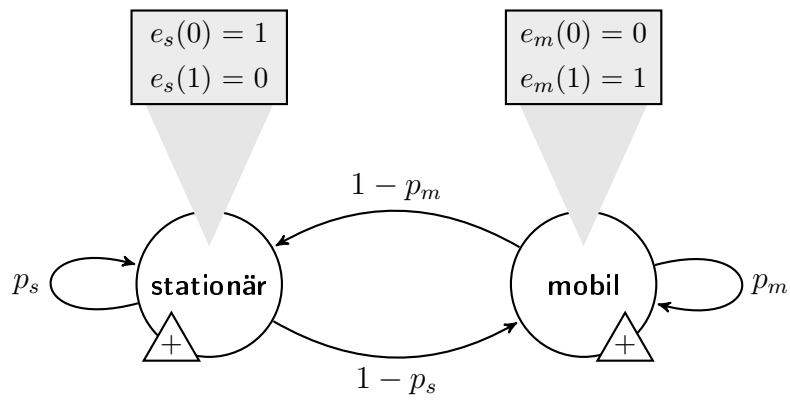


Abbildung 3.1: PAA für das 2-Parameter Modell

Kapitel 4

Implementierung

4.1 Simulation

4.1.1 Step-by-step

Es kann entweder jeder Zeitschritt der Chromatographie simuliert und dabei für jedes Teilchen entschieden werden, ob es sich bewegt. Dabei angenommen, dass die Wahrscheinlichkeit der Wechselwirkung unabhängig vom aktuellen Ort oder dem Zeitpunkt des letzten Phasenwechsels ist. Es wird für jedes Teilchen festgehalten, wo es sich befindet und ähnlich wie bei der echten Chromatographie beobachtet, wann wie viele Teilchen die Säule durchquert haben.

4.1.2 By-Event

Bei Übergangswahrscheinlichkeiten, die sehr groß oder klein sind, kommt es selten zu Zustandsänderungen, sodass es in diesem Fall effizienter ist, eine Wartezeitmethode zu nutzen: Es wird für jedes Teilchen entschieden, wann es den Zustand wechselt und, falls es mobil ist, wie weit es bis dahin weiter wandert. Es werden also nur die tatsächlichen Ereignisse simuliert, Schritte für Zeitpunkte, zu denen nichts passiert, entfallen.

4.2 PAA

Die Berechnung des PAA in Java, dabei Referenz auf mosdi (Zitiere Paper, in dem das vorgestellt wird und sage, dass ich das verwende)

Kapitel 5

Evaluation

Kapitel 6

Diskussion

Anhang A

Weitere Informationen

Abbildungsverzeichnis

3.1 PAA für das 2-Parameter Modell	6
--	---

Algorithmenverzeichnis

Literaturverzeichnis

- [1] KOLB, BRUNO: *Gaschromatographie in Bildern [eine Einführung]*. Wiley-VCH, Weinheim, 2., überarbeitete und erweiterte Auflage, 2003.
- [2] MARSCHALL, TOBIAS, INKE HERMS, HANS-MICHAEL KALTENBACH und SVEN RAHMANN: *Probabilistic Arithmetic Automata and Their Applications*. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 9(6):1737–1750, November 2012.