

Ein Gütemaß für Virtuelle Stochastische Sensoren - Rekonstruktionsgüte Typ-3

Pascal Krenckel*, Claudia Krull

Institut für Simulation und Graphik, Otto-von-Guericke-Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Deutschland; *krenckel@ovgu.de

Abstract. Virtuelle Stochastische Sensoren wurden an der Otto-von-Guericke-Universität Magdeburg entwickelt, um partiell-beobachtbare diskrete stochastische Modelle zu analysieren und das Systemverhalten zu rekonstruieren. Dabei existiert noch keine Beschreibung der Güte eines Virtuellen Stochastischen Sensors. In diesem Paper stellen wir ein Verfahren zur Bestimmung der Rekonstruktionsgüte vor. Dabei wird für zwei Modelle die Güte bestimmt. Diese Modelle wurden zusätzlich verändert, und die Auswirkung auf die Güte mit der erwarteten Güte verglichen. Unsere Ergebnisse zeigen, dass schon nach wenigen Zeitschritten ein Unterschied zwischen der Güte der verschiedenen Versionen der Modelle zu erkennen ist. Mit einem Gütemaß ist es möglich, verschiedene Sensorsetups zu vergleichen und diese zu optimieren.

Einführung

Virtuelle Stochastische Sensoren (VSS) können benutzt werden, um das Verhalten von partiell-beobachtbaren diskreten stochastischen Systemen zu rekonstruieren. Ähnlich zu den Virtuellen Sensoren [1, 2] messen VSS Größen, die nicht direkt messbar sind. Zum Beispiel kann ein Beschleunigungssensor benutzt werden, um die Geschwindigkeit abzuleiten. Auch ein VSS leitet das Systemverhalten aus einer gegebenen Sequenz von Beobachtungen ab – eine Zeitreihe physischer Sensordaten. Im Gegensatz zu Virtuellen Sensoren sind die Beziehungen zwischen Systemverhalten und Beobachtungen bei VSS nicht analytischer, sondern stochastischer Natur. Ein großes Problem ist dabei die Rekonstruktionsqualität des VSS. In diesem Paper stellen wir eine neue Definition der Rekonstruktionsgüte Typ-3 vor - ein Maß der Rekonstruktionsgüte eines VSS für die Pfadrekonstruktion. Durch die Rekonstruktionsgüte kann nicht nur die Anwendbarkeit eines VSS bestimmt werden, sondern auch verschiedene Sensorsetups verglichen werden. Dies ermöglicht es, das Sensorsetup zu optimieren.

1 Hintergrund

1.1 Virtuelle Stochastische Sensoren

VSS wurden 2011 eingeführt. Ein VSS modelliert einen doppelt stochastischen Prozess, bei dem nicht nur der interne Systemzustand von einem stochastischen Prozess abhängt, sondern auch die Ausgabe der „physischen Sensoren“. Die häufig verwendeten Hidden Markov Modelle (HMM) sind auch VSS [3, 4]. Es können aber auch Modelltypen mit gedächtnisbehaftetem Verhalten wie Augmented Stochastic Petri Nets (ASPN) verwendet werden. Ein VSS besteht immer aus einem Modell und einem Lösungsverfahren. Dabei gibt es drei Zielstellungen: Evaluierung, Decoding und Training. [3, 5]

Das Evaluierungsproblem versucht die Frage zu beantworten, wie wahrscheinlich bestimmte Ereignisse sind - z.B. eine gegebene Beobachtungssequenz. Dadurch kann entschieden werden, welches System wahrscheinlich eine bestimmte Beobachtungssequenz erzeugt hat.

Das Ziel des Decodings ist es, bei einer gegebenen Sequenz von Beobachtungen (Spur) die Sequenz der internen Zustände (Pfad) zu ermitteln. Dabei gibt es verschiedene Herangehensweisen. Die drei meist genutzten sind:

1. Der wahrscheinlichste Pfad
2. Die Sequenz der wahrscheinlichsten Zustände
3. Die Sequenz der wahrscheinlichsten Zustände, die immer noch ein Pfad ist.

Bei Hidden Markov Modellen heißen die entsprechenden Algorithmen, die diese Sequenzen berechnen: Viterbi, Posterior und Posterior-Viterbi. [4]

Beim Training wird versucht, das Parameterset eines Modells zu finden, das die höchste Wahrscheinlichkeit hat eine gegebene Menge von Beobachtungen zu erzeugen. Für gedächtnisbehaftete Modelle mit diskreten Zustandsräumen existieren bereits Lösungsverfahren die das Evaluierungsproblem und das Decodingproblem lösen können. Die Proxelmethode von Graham Horton (2002) und

deren Erweiterungen ermöglichen das Lösen, ohne dass Differenzialgleichungen gelöst werden müssen. Dafür wird die Zeitdimension diskretisiert und in Zeitschritten gerechnet. [6, 7, 8, 9]

1.2 Beobachtbarkeit

Der Begriff Beobachtbarkeit wurde von Kálmán im Kontext von linear dynamischen Systemen eingeführt. Beobachtbarkeit beschreibt dort, ob sich der Systemzustand rekonstruieren lässt. Ist der Systemzustand bekannt, können für jeden nachfolgenden Zeitpunkt die Zustandsvariablen bestimmt werden. Hat der Ausgabevektor weniger Dimensionen als der Eingabevektor, so wird mehr als ein Messwert benötigt, um den Systemzustand zu rekonstruieren. [10, 11, 12]

Für VSS ist diese Definition nicht geeignet. Der Begriff der Beobachtbarkeit soll im Kontext der VSS ein Maß für die Rekonstruktionsgüte des VSS sein und nicht nur ein binäres Maß. In [13] wurde erstmals die Beobachtbarkeit und damit ein Maß für die Güte eines VSS definiert. In früheren Publikationen haben wir den Begriff Beobachtbarkeit verwendet. Um Verwechslungen mit dem binären Maß aus der Regelungstechnik zu vermeiden, haben wir uns dazu entschieden, den Begriff in Rekonstruktionsgüte umzubenennen.

Im normalen Anwendungsfall ist die Ground-Truth, die echte Sequenz der internen Zustände, nicht bekannt. Die Aufgabe des VSS ist es, genau diese Sequenz zu berechnen. Im praktischen Einsatz, wäre es hilfreich, wenn insbesondere vor der Installation des Sensors bestimmt werden kann, welche Qualität der Sensor haben wird. Genau zur Lösung dieses Problems wurde die Rekonstruktionsgüte im Kontext der Virtuellen Stochastischen Sensoren eingeführt. Die Rekonstruktionsgüte wurde dabei in drei Typen unterteilt, die in den folgenden Abschnitten kurz erläutert werden. Sowohl in den vorangegangenen Arbeiten als auch für diese Arbeit wird angenommen, dass sowohl die Zeit als auch der Zustandsraum diskret sind. Ist dies nicht der Fall, so muss das Modell vorher diskretisiert werden. [13, 14]

Detektierbarkeit. Die Rekonstruktionsgüte Typ-1 beschreibt mit welcher Sicherheit ein auftretendes Ereignis beobachtet werden kann. Art und Eindeutigkeit werden dabei nicht betrachtet. Dadurch ist die Rekonstruktionsgüte Typ-1 leicht zu berechnen. Allerdings schränkt dies die Aussagekraft auch stark ein. So ist es in seltenen Fällen sogar möglich, dass zusätzliche Ausgaben die Rekonstruktionsgüte erhöhen, aber die Güte des Virtuellen Stochastischen Sensors senken.

Eine Typ-1 Rekonstruktionsgüte von 100% hat dennoch eine besondere Bedeutung. In diesem Fall ist jeder Zustandswechsel detektierbar. Dadurch ist ein Event-basierter Ansatz zum Lösen möglich. Die von Robert Buchholz entwickelten CHMMs sind genau so definiert, und ermöglichen die Verhaltensrekonstruktion partiell beobachtbarer diskreter stochastischer Systeme mit kontinuierlichem Zeitverlauf. Diese Methode ist um ein vielfaches schneller und genauer als das allgemeine VSS Lösungsverfahren. [13, 14, 15]

Zustandsrekonstruktion. Die Rekonstruktionsgüte Typ-2 löst das Problem der schlechten Aussagekraft der Rekonstruktionsgüte Typ-1. Die Rekonstruktionsgüte Typ-2 ist ein direktes Maß für die Güte der Rekonstruktion des internen Zustandes und ist definiert über die Summe aller Quadrate der Zustandswahrscheinlichkeiten. Die aktuelle Definition nutzt keine Ähnlichkeitsmaße, Zustände sind entweder identisch oder nicht. Eine exakte Berechnung der Rekonstruktionsgüte Typ-2 ist allerdings in der Praxis nicht möglich. Sowohl der Zustandsraum als auch die Anzahl der Spuren steigen exponentiell mit der Anzahl der Zeitschritte an. [13]

Die Rekonstruktionsgüte Typ-2 ist ein gutes Maß für die Güte der Zustandsrekonstruktion. Ist allerdings eine Sequenz von Zuständen oder Ereignissen von Interesse, dann ist die Rekonstruktionsgüte Typ-2 nicht anwendbar.

Pfadrekonstruktion. Da die Rekonstruktionsgüte Typ-2 für Aussagen über die Rekonstruktionsgüte der Zustands- oder Ereignissequenzen nicht anwendbar ist, wurde die Rekonstruktionsgüte Typ-3 eingeführt. Die in [13] vorgestellte Definition beschreibt die Rekonstruktionsgüte Typ-3 als erwartete Distanz zwischen allen Pfaden. Dafür werden entsprechende Distanzmaße benötigt. Welche dies sind, ist stark problemabhängig. Sowohl in [13] als auch in dieser Arbeit beschränken wir uns auf die Hamming-Distanz und die Damerau-Levenshtein-Distanz. Vorteil der Verwendung von Distanzmaßen ist, dass dadurch Ähnlichkeiten abgebildet werden können. Kleine Unterschiede in den Pfaden können so auch kleine Unterschiede in der Distanz zur Folge haben.

Allerdings gilt auch hier, dass in der Praxis eine exakte Berechnung der Rekonstruktionsgüte Typ-3 nicht möglich ist. Sowohl die Anzahl der Pfade, als auch die Anzahl der Spuren wächst exponentiell mit der Anzahl der Zeitschritte an. Schon nach wenigen Zeitschritten ist die Berechenbarkeitsgrenze erreicht. Eine Abschätzung mit der Monte-Carlo-Methode ist nicht möglich, da es nicht trivial ist, zwei Pfade mit gleicher Spur zu sampeln.

Zudem ist die Rekonstruktionsgüte Typ-3 über die Differenz aller Pfade zueinander definiert. In der Realität ist allerdings nur die Distanz zwischen dem Ground-Truth Pfad und dem berechneten Pfad von Interesse. Die Distanz zwischen zwei unwahrscheinlichen Pfaden ist nicht von Bedeutung. [6, 13, 14] Aus diesem Grund stellen wir hier eine neue Definition der Rekonstruktionsgüte Typ-3 vor, die nicht nur relevanter für den letztendlichen Anwendungsfall ist, sondern auch durch die Monte-Carlo-Methode abgeschätzt werden kann.

1.3 Distanzmaße für Sequenzen

Die in diesem Paper verwendeten Distanzmaße sind die Hamming-Distanz und die Damerau-Levenshtein-Distanz. Die Hamming-Distanz ist ein weit verbreitetes Distanzmaß und entspricht der Anzahl der Unterschiede zwischen zwei Zeichenketten. Dabei werden nur Zeichen mit gleichem Index miteinander verglichen. Dies hat zur Folge, dass nur Zeichenketten mit gleicher Länge sinnvoll verglichen werden können. Vorteil der Hamming-Distanz ist, dass sie sehr einfach und effizient zu berechnen ist.

Die Damerau-Levenshtein-Distanz ist die Anzahl der Operationen, die es benötigt, um die Zeichenketten ineinander umzuformen. Die Operationen sind: Einfügen, Ersetzen, Löschen eines Zeichens, sowie das Tauschen zweier benachbarter Zeichen. Die kleinste benötigte Anzahl an Operationen entspricht der Distanz. [16, 17] Werden nur Ersetzungsoperationen verwendet, so ist die Damerau-Levenshtein-Distanz gleich der Hamming-Distanz. Aber im Gegensatz zur Hamming-Distanz ermöglicht die Definition als Minimierungsproblem die Verwendung von Zeichenketten unterschiedlicher Länge. Zusätzlich kann das Löschen und Einfügen von Zeichen die Distanz stark senken. Abbildung 1 zeigt diesen Fall.

A	B	A	A	B	A	B	A	B	B	B	A	B	A	A	A	B	A	-
A	B	A	A	B	A	B	A	B	A	B	A	B	A	A	A	B	A	A

Abb. 1: Zwei Zeichenketten mit Damerau-Levenshtein-Änderung (rot) und Hamming-Abweichung (gelb)

Die Damerau-Levenshtein Distanz in Abbildung 1 ist eins, da lediglich das hervorgehobene „A“ hinzugefügt bzw. gelöscht werden muss. Die Hamming-Distanz ist hingegen sieben. Dabei wurde eine angepasste Version der Hamming-Distanz verwendet, die „Null“-Zeichen am Ende hinzufügt. Dadurch haben beide Zeichenketten die gleiche Länge. Insbesondere bei der Verwendung von

VSS und der Berechnung von Ereignissequenzen treten solche Unterschiede sehr häufig auf.

Ein Nachteil der Damerau-Levenshtein-Distanz ist, dass die Laufzeit der Berechnung in $O(n^2)$ liegt. Zudem sind beide Distanz-Maße nicht direkt vergleichbar, da es auf den Anwendungsfall oder die Fragestellung ankommt.

2 Neu Definition der Rekonstruktionsgüte Typ-3

Die Rekonstruktionsgüte Typ-3 versucht die Qualität des rekonstruierten Pfades eines VSS zu quantifizieren, vor dem tatsächlichen Einsatz. In [13] wurde die Rekonstruktionsgüte Typ-3 als erwartete Distanz zwischen allen möglichen Pfaden definiert. Die neu definierte Rekonstruktionsgüte Typ-3 ist die erwartete Distanz zwischen allen Pfaden und dem rekonstruierten Pfad.

Damit die Rekonstruktionsgüte Typ-3 in dem Bereich von $[0;1]$ liegt wird die berechnete Distanz zweier Pfade zusätzlich durch einen Normierungsfaktor geteilt und die Differenz zu 1 gebildet. Für die Hamming und Damerau-Levenshtein-Distanz entspricht der Normierungsfaktor der Pfadlänge.

$$O_3(t) = 1 - \frac{E(\text{Distanz}_{\text{Decoding}})}{N(t)}$$

$E(\text{Distanz}_{\text{Decoding}})$ ist die erwartete Distanz zwischen Ground-Truth und dem decodierten Pfad, also der erwarteten Distanz zwischen der Realität und dem VSS Output. Daher ist die neue Definition semantisch passender als die bisherige.

Der Erwartungswert $E(\text{Distanz}_{\text{Decoding}})$ hängt von zwei Zufallsprozessen ab, dem äußeren (den Systemausgaben) und dem inneren (der Systementwicklung). Dabei ist der decodierte Pfad von der Spur abhängig.

$$\begin{aligned} E(\text{Distanz}_{\text{Decoding}}) &= E\left(E(\text{Distanz}_{\text{Decoding}} | \text{Spur})\right) \\ &= E\left(\frac{P(\text{Distanz}_{\text{Decoding}}, \text{Spur})}{P(\text{Spur})} \times \text{Distanz}_{\text{Decoding}}\right) \\ &= E\left(P(\text{Distanz}_{\text{Decoding}} | \text{Spur}) \times \text{Distanz}_{\text{Decoding}}\right) \\ &= \sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(s) \times P(p|s) \times \text{Distanz}(p, \text{dec}(s)) \end{aligned}$$

$\text{Distanz}(\dots)$ berechnet die Distanz zwischen dem Pfad p und dem decodierten Pfad $\text{dec}(s)$ aus Spur s . $P(s)$

ist die Wahrscheinlichkeit die Spur s zu beobachten und $P(p|s)$ ist die Wahrscheinlichkeit des Pfades p unter der Bedingung, dass s beobachten wurde. Die Pfadlänge $|p|$ ist dabei konstant - die Anzahl der Zeitschritte $t + 1$. Damit ergibt sich für die Rekonstruktionsgüte:

$$O_3(t) = 1 - \frac{\sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(s) \times P(p|s) \times \text{Distanz}(p, \text{dec}(s))}{N(t)}$$

Mit einem sinnvollen Normierungsfaktor ergeben sich folgende Grenzwerte:

- $O_3(t) = 1$
Die Distanz zwischen den Pfaden ist 0. Der Pfad ist bezogen auf das Distanzmaß vollständig rekonstruierbar.
- $O_3(t) = 0$
Die Distanz zwischen den Pfaden ist maximal. Der Pfad ist nicht rekonstruierbar. Dieser Wert ist jedoch nicht zu erreichen, da die Distanz des Pfades zu sich selber Null ist.

Die Anzahl der Pfade wächst exponentiell. Im Gegensatz zur vorherigen Definition wird aber jeder Pfad nur mit dem decodierten Pfad verglichen. Dies reduziert den Rechenaufwand deutlich. Dennoch ist auch dies in der Praxis nicht berechenbar.

2.1 Abschätzung

Die Distanz kann allerdings durch einen Maximum-Likelihood-Schätzer sehr gut abgeschätzt werden. Mithilfe einer Monte-Carlo Simulation werden verschiedene Pfade mit entsprechender Spur erzeugt. Aus der Spur wird mithilfe des Decodings ein entsprechender Pfad rekonstruiert und die Distanz zum richtigen Pfad bestimmt. Durch Anwenden des Satz von Bayes erhalten wir eine Formel, die durch die Monte-Carlo Simulation abgeschätzt werden kann:

$$\begin{aligned} & \sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(s) \times P(p|s) \times \text{Distanz}(p, \text{dec}(s)) \\ &= \sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(s) \times \frac{P(p \wedge s)}{P(s)} \times \text{Distanz}(p, \text{dec}(s)) \\ &= \sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(p \wedge s) \times \text{Distanz}(p, \text{dec}(s)) \\ &= \sum_{s \in \text{Spur}} \sum_{p \in \text{Pfad}} P(p) \times P(s|p) \times \text{Distanz}(p, \text{dec}(s)) \end{aligned}$$

Das Sampeln des Pfades p mit der entsprechenden Wahrscheinlichkeit $P(p)$ ist per Konstruktion möglich. Das Sampeln einer Spur s für einen Pfad ist ebenfalls tri-

vial möglich. Diese Abschätzung ist zum einen erwartungstreu, zum anderen kann durch die Streuung des Datensatzes ein Konfidenzintervall berechnet werden.

Im folgenden Experimente Kapitel zeigen wir, dass die Abschätzung berechenbar und die Ungenauigkeit sehr klein ist.

3 Experimente

Für die Experimente wurden zwei einfache Modelle mit unterschiedlicher Zielsetzung verwendet. Diese Modelle werden in verschiedenen Varianten verwendet, deren Systemzustand je nach Setup, unterschiedlich gut rekonstruierbar sein sollten.

Es wurden drei verschiedene Decoding-Verfahren, sowie die Hamming-Distanz und Damerau-Levenshtein-Distanz als Distanzmaß verwendet. Dabei wurde ein Pfad einmal als Sequenz von internen Zuständen und einmal als Sequenz von Ereignissen betrachtet. Zur Abschätzung der Rekonstruktionsgüte wurden alle Experimente mit wachsender Pfadlänge von bis zu 100.000 simuliert. Für Pfade mit einer Länge von bis zu 1.000 wurden die System je Pfadlänge 1.000-mal simuliert. Pfade mit einer Länge von über 1.000 wurden hingegen 100-mal simuliert. Jede Simulation entspricht einem zufälligen Ground-Truth Pfad mit entsprechender Beobachtungssequenz, aus der ein Pfad rekonstruiert werden kann. Um die Aussagekraft der Rekonstruktionsgüte Typ-3 zu testen, haben wir jedes Modell manipuliert und so die Rekonstruktionsgüte verändert.

3.1 Car-Rental-Service

Das erste System stellt den Bedienprozess einer Autovermietung dar. Die Autovermietung hat im Grundsystem eine Tür. Alle Kunden betreten und verlassen die Autovermietung durch diese Tür. Passiert ein Kunde die Tür, so wird dies über einen Sensor registriert. Es gibt Premium-Kunden, die mit Priorität behandelt werden, und eine längere Bedienzeit haben. Es gibt einen Angestellten. Ziel des VSS ist es, die Länge der Warteschlangen anhand des Türprotokolls zu bestimmen.

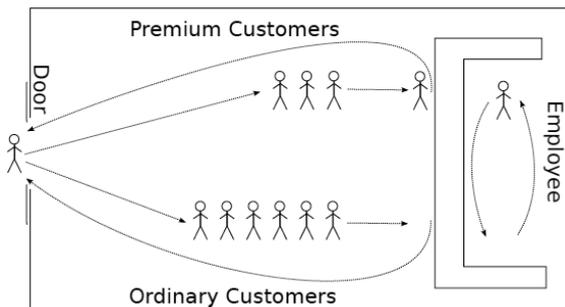


Abb. 2: Autovermietung mit zwei getrennten Warteschlangen für Premium und normale Kunden und einem Angestellten

Die mittlere Bearbeitungszeit pro Kunde (11.8 Zeiteinheiten) ist geringer als die mittlere Zwischenankunftszeit (15 Zeiteinheiten). Dies ist wichtig, da sich sonst, die Kunden aufstauen würden. Da das Ziel die Bestimmung der Warteschlangenlänge ist, wird hier der Pfad als Sequenz von internen Zuständen betrachtet. Der Diskretisierungszeitschritt im Lösungsverfahren beträgt 5 Zeiteinheiten. Im Schnitt tritt damit in 2/3 aller Zeitschritte ein Ereignis ein. Um die berechnete Rekonstruktionsgüte mit der „gefühlten“ Rekonstruktionsgüte zu vergleichen, wurden vier Varianten dieses Modells erzeugt.

1. Im Grundmodell existiert eine Tür mit einer Lichtschranke. Jedes Passieren der Tür erzeugt dasselbe Ausgabesymbol. (*Original*)
2. In der zweiten Version kann durch eine doppelte Lichtschranke die Richtung der Personen bestimmt werden. (*Double*)
3. In der dritten Version gibt es einen extra Eingang nur für Premiumkunden. 30% der Premiumkunden benutzen diesen. Ein Premiumkunde muss nicht durch dieselbe Tür ein und austreten. (*VIP*)
4. In der vierten Version wird zwischen Eingang und Ausgang und Normal- und Premiumkunde unterschieden. Jedes Ereignis ist exakt beobachtbar. (*Komplett*)

Die vier verschiedenen Modelle wurden verwendet, um die berechnete Rekonstruktionsgüte mit der gefühlten Rekonstruktionsgüte zu vergleichen. Für die gefühlte Rekonstruktionsgüte ergibt sich folgendes Ranking: Variante 1 hat die geringste Rekonstruktionsgüte, da zwar alle Ereignisse detektierbar sind, aber weder zwischen Kunden, noch zwischen Betreten und Verlassen unterschieden werden kann. Bei Version zwei (Doppelte Lichtschranke) ist das Betreten und Verlassen eindeutig zu unterscheiden. Premiumkunden betreten die Autovermietung deutlich seltener (1:2). Die doppelte Lichtschranke (Version 2) sollte also eine deutliche Verbesserung der

Rekonstruktionsgüte zur Folge haben. Version 3 hingegen sollte die Rekonstruktionsgüte nur leicht verbessern, da nur wenig Events besser zu beobachten sind. Variante vier ist zu 100% beobachtbar. Daher ist folgende Rangfolge zu erwarten: 4, 2, 3, 1.

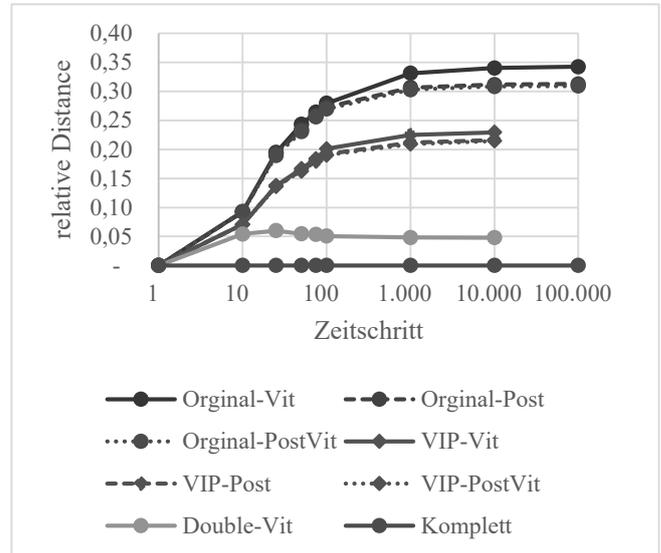


Diagramm 1: Relative Hamming Distanz für Viterbi, Posterior und Posterior-Viterbi für alle 4 Varianten

Diagramm 1 zeigt den Verlauf der relativen Distanz über die Pfadlänge für die Hamming-Distanz des Posterior-Viterbi-Algorithmus. Da die Werte im Bereich von bis zu 0,3 liegen, ist hier, zur besseren Übersichtlichkeit, die relative Distanz und nicht deren Differenz zu 1 abgebildet. Daher ist ein größerer Wert in den Diagrammen schlechter. Außerdem ist die x-Achse in logarithmischer Skala eingeteilt. Gut zu erkennen ist, dass die Rekonstruktionsgüte konvergiert. Von 1.000 zu 10.000 Zeitschritten verändert sich der Wert der Rekonstruktionsgüte nur noch um 0.01. Von 10.000 zu 100.000 ist kein signifikanter Unterschied mehr zu sehen. Das 99%-Konfidenzintervall hat eine Breite von maximal ± 0.01 .

Posterior und Posterior-Viterbi Algorithmus liefern annähernd dieselben Ergebnisse. Der Viterbi-Algorithmus ist geringfügig schlechter. Deutlich ist zu erkennen, dass das Originalmodell die schlechtesten Werte liefert. Danach folgt das VIP-Modell (3) das leicht bessere Werte liefert. Das Modell mit der doppelten Lichtschranke (2) und das komplett-beobachtbare System (4) liefern deutlich bessere Ergebnisse. Dies entspricht der erwarteten Reihenfolge. Auch deutlich zu erkennen ist, dass bereits

nach wenigen Zeitschritten die Reihenfolge der verschiedenen Versionen bezüglich der Rekonstruktionsgüte sichtbar ist. Die Ergebnisse für die Damerau-Levenshtein-Distanz sind hier nicht abgebildet, da sie für dieses Model dieselben Werte liefert, wie die Hamming-Distanz. Dies liegt an der Art, wie die Pfade betrachtet werden. Ein Pfad entspricht dem internen Zustand des Systems. Werden zwei Events in der falschen Reihenfolge erkannt, so unterscheiden sich die Zustände nur nach dem ersten Event. Nach dem zweiten Event hätten beide Pfade denselben internen Zustand. Außerdem sind alle Ereignisse beobachtbar. Die zusätzlichen Operationen (Tauschen, Hinzufügen, Löschen) machen daher kaum einen Unterschied.

3.2 Quality-Tester

Das zweite System beschreibt einen Quality-Tester. Das System besteht aus zwei Quellprozessen (Source 0 und Source 1), die mit unterschiedlicher Häufigkeit Items produzieren. Die Items werden dann einem Qualitätstest unterzogen. Der Tester gibt dann für jedes Item den Zeitstempel und den Zustand des Items aus (OK oder NOK). Dabei ist nicht bekannt von welcher Quelle das Item stammt. Ziel ist es, zuzuordnen, welche Quelle welches Item produziert hat. In unserem Fall ist die Art des Ausgabesymbols nicht wichtig, da beide Quellen mit gleicher Wahrscheinlichkeit defekte Items produzieren. Abb. 3 zeigt den Quality-Tester.

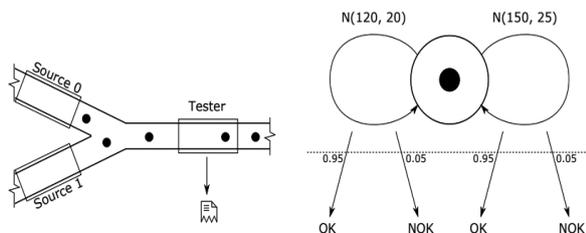


Abb. 3: Systemabbildung und Zustandsraum Model des Quality-Testers

Interessant ist dieses Modell, da es mit nur zwei Transitionen auskommt. Alle internen Zustände sind dieselben. Daher muss der Pfad hier als Sequenz von Ereignissen betrachtet werden. Aus diesem Grund werden hier auch nur die Daten auf Basis des Viterbi-Decodings vorgestellt. Posterior und Posterior-Viterbi sind nicht geeignet, um Event-Sequenzen zuverlässig zu bestimmen. Dies liegt daran, dass sie nur die Wahrscheinlichkeiten der einzelnen Zustände betrachten. Die Zustandsübergänge, die die Events darstellen, spielen bei Posterior keine

Rolle. Um die Rekonstruktionsgüte zu reduzieren, wurde das Model angepasst. Jedes Item wird nur mit einer 50% Wahrscheinlichkeit getestet (Half).

Komplette Ereignissequenz. Da im Grundmodell alle Ereignisse beobachtbar sind, werden alle Zeitschritte ohne Event mit einem Null-Event aufgefüllt. Ein Null-Event bedeutet, dass kein Ereignis eingetreten ist. Da Null-Events den Großteil des Pfades ausmachen, ist die relative Distanz auch entsprechend klein. Zudem zeigt sich auch in diesem Experiment, dass Hamming-Distanz und Damerau-Levenshtein-Distanz ähnliche Ergebnisse liefern.

Anders verhält es sich aber mit der Manipulierten Variante. Da nur 50% der Ereignisse beobachtbar sind, muss hier auch der Zeitschritt, in dem das Ereignis auftrat, korrekt rekonstruiert werden. Dadurch steigt die relative Distanz stark an. Hier zeigt sich auch ein Unterschied zwischen Hamming- und Damerau-Levenshtein-Distanz. Diagramm 2 zeigt die relativen Distanzen.

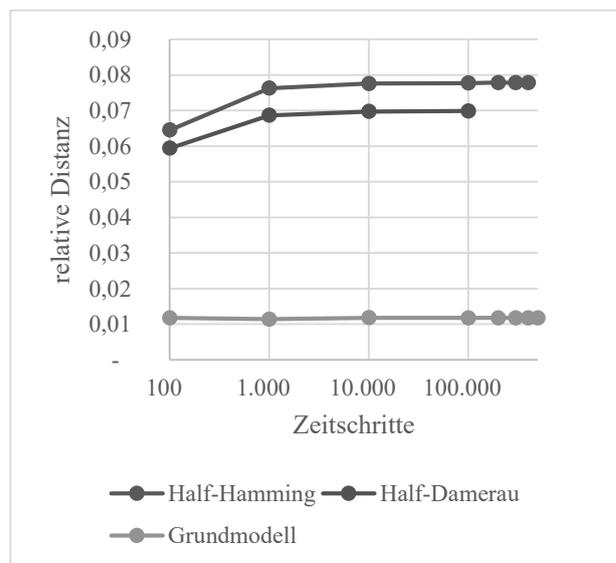


Diagramm 2: relative Distanz für das Grundsystem und dem System mit reduzierten Ausgaben mit dem Viterbi-Algorithmus.

Für das Grundmodell erzielt die Damerau-Levenshtein-Distanz nur im Falle, dass beide Quellen in aufeinanderfolgenden Zeitschritten ein Item produzieren und diese falsch zugeordnet wurden, bessere Ergebnisse. Dann ist die Hamming-Distanz für diesen Sequenzabschnitt zwei, während die Damerau-Levenshtein-Distanz eins ergibt. Diese Fälle sind im Grundmodell sehr selten. Im Modell mit reduzierter Rekonstruktionsgüte sieht das allerdings anders aus. Wird ein Item das keine Ausgabe produziert

einem benachbarten Zeitschritt zugeordnet, so ist auch hier die Damerau-Levenshtein Distanz kleiner. Die Ergebnisse zeigen, dass die Rekonstruktionsgüte schnell konvergiert und dass auch rechnerisch das Modell mit reduzierter Rekonstruktionsgüte schlechtere Ergebnisse liefert. Allerdings zeigt sich hier, dass die Werte selber von geringer Aussagekraft sind, da hauptsächlich Null-Events auf Null-Events zugeordnet wurden, was bei den verwendeten Modellen keine große Schwierigkeit darstellt, da fast alle Events Null-Events sind.

Eventreihenfolge. Um das Problem der Null-Events zu lösen, wurde im zweiten Experiment mit den beiden Modellen, der Pfad nur auf die Events ungleich dem Null-Event eingekürzt. Dadurch werden die Pfade deutlich kürzer.

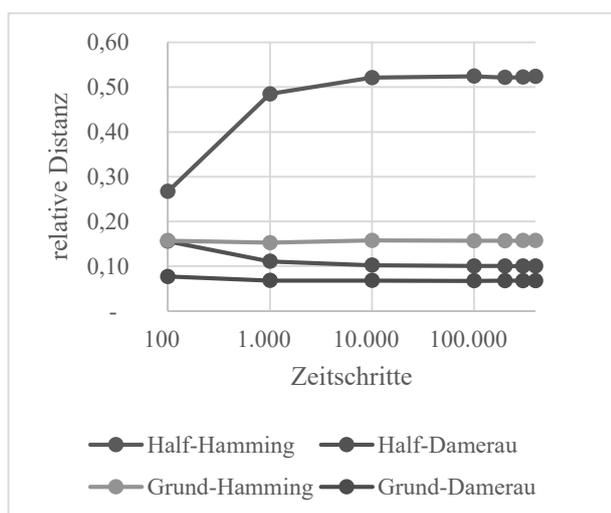


Diagramm 3: relative Distanz des Quality-Testers der verschiedenen Systeme mit Damerau- und Hamming-Distanz des Experiments „Eventreihenfolge“

Auch hier zeigt sich, dass die Rekonstruktionsgüte schnell konvergiert. Zusätzlich zeigt sich hier, dass die Damerau-Levenshtein-Distanz deutlich besser performt als die Hamming-Distanz. Insbesondere ist die relative Distanz des Modells mit reduzierter Rekonstruktionsgüte $>50\%$. Dies liegt daran, dass die Hamming-Distanz eigentlich für dieses Problem nicht verwendet werden kann. Da nicht jedes Ereignis beobachtet wird, ist auch die Anzahl der produzierten Items nicht bekannt. Es entstehen Pfade mit unterschiedlicher Länge. Damit die Hamming-Distanz damit umgehen kann, wurde der längere Pfad abgeschnitten und die abgeschnittene Länge der Distanz zugerechnet.

Bei dem Grundmodell kann ein derartiges Problem

nicht auftreten, da jedes Ereignis beobachtet wird. Wird allerdings ein Ereignis falsch erkannt, so bedeutet dies in fast allen Fällen, dass das darauffolgende Ereignis mit hoher Wahrscheinlichkeit auch falsch erkannt wurde.

4 Ergebnisdiskussion

Für die verwendeten Modelle war es möglich, die Rekonstruktionsgüte selbst für viele Zeitschritte zu berechnen. Es zeigt sich, dass insbesondere Posterior und Posterior-Viterbi sehr gut bei der Rekonstruktion der internen Zustände funktionieren. Allerdings sind diese Algorithmen bei der Rekonstruktion des Event-Pfads nicht verwendbar. Dies ist zu erwarten, da insbesondere Posterior keine Pfade, sondern nur Zustandssequenzen zurückgibt.

Auch zeigt sich, dass die Hamming-Distanz nicht sinnvoll angewendet werden kann, wenn die Ereignisse nicht zu 100% zu beobachten sind.

Zu beobachten ist auch, dass bei allen Experimenten die Rekonstruktionsgüte Typ-3 konvergiert. Mit größerem zeitlichem Abstand wird der Einfluss verschiedener Systemzustände immer geringer und nähert sich 0 an.

Der Unterschied in der Rekonstruktionsgüte Typ-3 war bereits nach wenigen Zeitschritten zu erkennen. Dies ist eine wichtige Entdeckung. Für die hier verwendeten Modelle ist ein Berechnen der Rekonstruktionsgüte Typ-3 für lange Sequenzen nicht notwendig, wenn nur die Reihenfolge der Güte der VSS interessant ist. So lässt sich das Sensorsetup optimieren und das beste Sensorsetup bestimmen.

Einschränkend ist allerdings zu sagen, dass bei der Rekonstruktion des Pfades für die sehr langen Spuren einzelne Pfade nicht rekonstruiert werden konnten. Dies liegt an der verwendeten Methode zur Rekonstruktion. Auch wenn diese Fälle sehr selten auftraten, so sind dies grade die unwahrscheinlichen Pfade, bei denen eine hohe Distanz zum dekodierten Pfad zu erwarten ist. Dadurch wird die Distanz geringfügig unterschätzt.

Ein hier nicht untersuchter Einflussfaktor ist die Länge eines Zeitschrittes. Je kürzer die Zeitschritte, desto genauer, aber auch rechenaufwändiger sind die Ergebnisse. Der in diesem Paper verwendete Algorithmus setzt voraus, dass nur ein Ereignis im selben Zeitschritt auftritt. Insbesondere bei der Autovermietung treten alle Ereignisse mit einer Dichte von $1/2$ bis zu $2/3$ auf. Dies betrifft allerdings auch die letztendliche Verwendung des VSS. Auch dort besteht dasselbe Problem, sodass die Aussagen zur Rekonstruktionsgüte dennoch korrekt sind.

Wird ein anderer Zeitschritt zur Lösung eines VSS verwendet, dann muss dies allerdings auch bei der Berechnung der Rekonstruktionsgüte des VSS berücksichtigt werden.

5 Zusammenfassung und Ausblick

Die Rekonstruktionsgüte ist ein wichtiges Werkzeug, um den potentiellen Nutzen eines Virtuellen Stochastischen Sensors zu bestimmen. Wir haben in diesem Paper eine Definition vorgestellt, die Bezug zur Verwendung hat. Unsere Experimente haben gezeigt, dass die Rekonstruktionsgüte Typ-3 nicht nur berechenbar ist, sondern bereits wenige Zeitschritte ausreichen, um die Modelle zu vergleichen. Dies ist notwendig, um das Sensorsetup zu optimieren, oder zu bestimmen ob ein Virtueller Stochastischer Sensor für einen bestimmten Anwendungsfall einsetzbar ist.

Aus dieser Arbeit ergeben sich allerdings auch einige weitere Fragen. So ist bisher noch nicht geklärt was eine Rekonstruktionsgüte mit einem spezifischen Wert genau für ein Problem bedeutet. Dies ist problemabhängig. Insbesondere die Distanzfunktion hat starken Einfluss auf die Rekonstruktionsgüte. Wie sich die Rekonstruktionsgüte mit anderen Distanzmaßen und größeren Modellen verhält, muss in zukünftigen Arbeiten noch untersucht werden. So könnte z.B. für die Autovermietung die mittlere Quadratische Abweichung von Interesse sein. Ob sich für ein derartiges Distanzmaß ein sinnvoller Normierungsfaktor findet, oder ob das Distanzmaß selber aussagekräftiger ist, ist auch eine Frage die noch geklärt werden muss.

6 Verweise

- [1] S. Kabadayi, A. Pridgen und C. Julien, „Virtual sensors: abstracting data from physical sensors,“ *2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks(WoWMoM'06)*, pp. 586-592, 2006.
- [2] L. Fortuna, S. Graziani, A. Rizzo und M. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*, 2007.
- [3] C. Krull, *Virtual Stochastic Sensors: Formal Background and Example Applications*, Magdeburg: Shaker Verlag Düren, 2021.
- [4] L. Rabiner und B. Juang, „An introduction to hidden Markov models,“ *IEEE ASSP Magazine*, Bd. Volume 3, Nr. 1, pp. 4-16, Jan 1986.
- [5] C. Krull, R. Buchholz und G. Horton, „Virtual Stochastic Sensors: How to gain Insight into Partially Observable Discrete Stochastic Systems,“ in *The 30th IASTED International Conference on Modelling, Identification and Control*, 2011.
- [6] G. Horton, „A NEW PARADIGM FOR THE NUMERICAL SIMULATION OF STOCHASTIC PETRI NETS WITH GENERAL FIRING TIMES,“ in *European Simulation Symposium*, Dresden, 2002.
- [7] S. Lazarova-Molnar und G. Horton, *Proxel-Based Simulation of Stochastic Petri Nets*, SCS European Publishing House, 2004.
- [8] S. Lazarova-Molnar und G. Horton, *Proxel-Based Simulation of Stochastic Petri Nets Containing Immediate Transitions*.
- [9] F. Wickborn, C. Isensee, T. Simon, S. Lazarova-Molnar und G. Horton, *A New Approach for Computing Conditional Probabilities of General Stochastic Processes*, 2006.
- [10] Kálmán, *LECTURES ON CONTROLLABILITY AND OBSERVABILITY*, Stanford University, 1970.
- [11] J. Lunze, *Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung*, Springer-Verlag Berlin Heidelberg GmbH, 2002.
- [12] F. M. Ham und R. G. Brown, „Observability, Eigenvalues, and Kalman Filtering,“ in *IEEE Transactions on Aerospace and Electronic Systems* 2, 1983.
- [13] P. Krenckel, *Beobachtbarkeit von Virtuellen Stochastischen Sensoren*, Magdeburg: Otto-von-Guericke-Universität Magdeburg, 2019.
- [14] P. Krenckel, C. Krull und G. Horton, *Observability of Virtual Stochastic Sensors: Observability Type-1*, Palma de Mallorca, 2019, pp. 8-12.
- [15] R. Buchholz, *Conversive Hidden non-Markovian models*, Magdeburg, 2012.
- [16] F. J. Damerau, „A technique for computer detection and correction of spelling errors,“ in *Communications of the ACM*, Bd. 7, New York, NY, USA, Association for Computing Machinery, 1964, pp. 171-176.
- [17] V. I. Levenshtein, „Binary Codes Capable of Correcting Deletions, Insertions and Reversals,“ *Soviet Physics Doklady*, Bd. 10, p. 707, February 1966.