

Maße für Wahrscheinlichkeitsdichten in der informationstheoretischen Sensoreinsatzplanung

Daniel Lyons, Achim Hekler, Benjamin Noack und Uwe D. Hanebeck

Lehrstuhl für Intelligente Sensor-Aktor-Systeme (ISAS),
Institut für Anthropomatik,
Karlsruher Institut für Technologie (KIT)
<http://isas.uka.de>

Zusammenfassung Bei der Beobachtung eines räumlich verteilten Phänomens mit einer Vielzahl von Sensoren ist die intelligente Auswahl von Messkonfigurationen aufgrund von beschränkten Rechen- und Kommunikationskapazitäten entscheidend für die Lebensdauer des Sensornetzes. Mit der Sensoreinsatzplanung kann die im nächsten Zeitschritt anzusteuern- de Messkonfiguration dynamisch mittels einer stochastischen modell-prädiktiven Planung über einen endlichen Zeithorizont bestimmt werden. Dabei wird als Gütekriterium die Maximierung des zu erwartenden Informationsgewinns durch zukünftige Messungen unter sparsamer Verwendung der Energieressourcen gewählt. In diesem Artikel wird ein neues Maß für kontinuierliche Wahrscheinlichkeitsdichten vorgestellt, das sich kanonisch aus der Konstruktion eines Vektorraums für Wahrscheinlichkeitsdichten ergibt. Dieses Maß wird als Gütefunktion in der vorausschauenden Sensoreinsatzplanung zur Bewertung des informationstheoretischen Einfluß von Messungen auf die aktuelle Zustandsschätzung verwendet.

1 Einleitung

Mit einem Sensor-Netzwerk werden Daten über ein verteiltes Phänomen gewonnen und diese dann in einem oder mehreren lokalen Verarbeitungszentren geeignet miteinander fusioniert. Um eine hochqualitative Zustandsschätzung des Phänomens zu erhalten, wäre der Einsatz von

möglichst vielen Sensoren wünschenswert. Dies ist aber in konkreten Anwendungen generell nicht realisierbar: einerseits kostet jede Messung eines Sensors Energie, andererseits entstehen durch die Übertragung des Messwerts über ein Kommunikationsnetzwerk hohe Energiekosten. Da in einem Sensor-Netzwerk die Energieressourcen der einzelnen Sensorknoten in der Regel beschränkt sind, ist es ratsam, dass nicht jeder Sensor zu jedem Zeitpunkt Messungen vornimmt, sondern nur diejenigen Sensoren für Messungen ausgewählt werden, deren Messungen auch einen hohen Informationsgehalt versprechen.

Diese Auswahl von Sensoren für zukünftige Messungen übernimmt ein sogenannter *Sensormanager*. Dieser muß hierbei auch berücksichtigen, dass die von dem Sensor erhaltene Information durch Messrauschen beeinflusst sein wird. Zunächst muss der Sensormanager dafür eine geeignete Gütefunktion wählen, die den zukünftigen Informationsgewinn bewertet, und dann vorausschauend eine optimale Sensorsequenz bestimmen, die den akkumulierten Informationsgewinn in Abhängigkeit vom aktuellen Zustand maximiert.

2 Problemformulierung

Das Ziel der Sensoreinsatzplanung ist es, den Zustand des beobachteten physikalischen Phänomens mit Hilfe eines Sensor-Netzwerks möglichst präzise unter Verwendung von möglichst wenigen Sensoren zu schätzen. Im Folgenden nehmen wir an, dass die räumliche und zeitliche Entwicklung des Phänomens durch eine zeitdiskrete und wertkontinuierliche Systemfunktion beschrieben wird, die den Zustand des Phänomens zum Zeitpunkt k auf den Zustand zum Zeitpunkt $k + 1$ abbildet.

2.1 Sensorauswahl

Bei einem naiven Vorgehen führt jeder Sensor zu jedem Zeitpunkt eine Messung durch, die dann geeignet miteinander funskioniert werden. Als Ergebnis erhält man zwar eine hochqualitative Zustandsschätzung, die Lebensdauer des Netzwerks würde aber aufgrund des dadurch entstehenden Energieverbrauchs durch Messungen und Kommunikation stark verkürzt werden. Daher ist es sinnvoll, ein Gleichgewicht zwischen einer guten Schätzqualität und einer Verlängerung der Lebensdauer des Netzwerks zu finden. Der oben erwähnte Sensormanager versucht den erwartete

ten Informationsgewinn $E\{R(\underline{x}_k, \underline{s}_k)\}$ über den Zustand \underline{x}_k des Systems zu maximieren, wobei die Funktion $R(\cdot)$ den zukünftigen Informationsgewinn bewertet und $\underline{s}_k = [s_k^1, \dots, s_k^M]$ eine Sequenz von Sensoren ist, die für Messungen ausgewählt wurde. Dieser erwartete zukünftige Informationsgewinn soll modell-prädiktiv, das heißt unter Berücksichtigung der zukünftigen Entwicklung des Phänomens, maximiert werden [1]. In Kapitel 5.1 wird der konkrete Formalismus unter Anwendung der dynamischen Programmierung erläutert [2]. Die Lösung dieses Optimierungsproblems ist eine Sequenz von Sensorkonfigurationen, deren zukünftige Messungen den erwarteten Informationsgewinn über einen bestimmten Zeithorizont maximieren. Die erste Sensorkonfiguration in dieser optimalen Sequenz wird an das Sensornetzwerk angelegt. Die dadurch erhaltenen Messungen werden, wie im nächsten Kapitel beschrieben, verarbeitet. Im nächsten Zeitschritt erfolgt dann eine erneute Planung durch den Sensormanager.

2.2 System- und Messmodelle

Wir gehen davon aus, dass die Entwicklung des Zustands des zu beobachtenden Phänomens durch folgende Modellgleichung gegeben ist

$$\underline{x}_{k+1} = \underline{a}_k(\underline{x}_k, \underline{w}_k) , \quad (11.1)$$

wobei alle Größen skalar oder vektorwertig sein können. Die möglicherweise zeitvariante Systemfunktion $\underline{a}_k(\cdot)$ bildet den Zustand \underline{x}_k zum Zeitpunkt k auf den Zustand \underline{x}_{k+1} zum Zeitpunkt $k + 1$ ab. Das Systemrauschen \underline{w}_k subsummiert dabei alle auftretenden Störungen, wie z.B. exogene Störeinflüsse oder Modellierungsfehler, und wird stochastisch modelliert. Im Folgenden repräsentieren alle fett geschriebenen Größen Zufallsvariablen. Ist $f_k(\underline{x}_k)$ die Wahrscheinlichkeitsdichte, die den aktuellen Zustand \underline{x}_k beschreibt, so kann mit Hilfe des Systemmodells (11.1) die Entwicklung des Systems mittels der unterlagernden Dichte prädiziert werden

$$f_{k+1}^P(\underline{x}_{k+1}) = \int_{\Omega} f_k^T(\underline{x}_{k+1}|\underline{x}_k) f_k(\underline{x}_k) d\underline{x}_k , \quad (11.2)$$

wobei f^T die Transitionsdichte ist und aus dem Modell (11.1) bestimmt werden kann.

Die prädizierte Zustandsdichte kann dann mit neu erhaltenen Messungen fusioniert werden. Dafür wird das Messmodell eines Sensors

$s_k \in \{1, \dots, L\}$ benötigt, das den Systemzustand mit dem Ausgang einer Messung in Zusammenhang setzt. Es wird durch eine nichtlineare Abbildung

$$\hat{\underline{z}}_{s_k} = \underline{h}_{s_k}(\underline{x}_k, \underline{v}_{s_k})$$

beschrieben, wobei $\hat{\underline{z}}_{s_k}$ der konkrete Ausgang der Messung und \underline{v}_{s_k} das stochastische Messrauschen ist, das für jeden Sensor verschieden sein kann. Die prädiizierte Dichte f_k^p wird mit der erhaltenen Meßinformation nach der Bayes Formel

$$f_{s_k}^e(\underline{x}_k) = \frac{f_{s_k}^L(\hat{\underline{z}}_{s_k}|\underline{x}_k)f_k^p(\underline{x}_k)}{\int_{\Omega} f_{s_k}^L(\hat{\underline{z}}_{s_k}|\underline{x}_k)f_k^p(\underline{x}_k)d\underline{x}_k} \quad (11.3)$$

fusioniert, wobei $f_{s_k}^L(\hat{\underline{z}}_{s_k}|\underline{x}_k)$ die sogenannte Likelihood darstellt.

3 Informationstheoretische Gütemaße

Um eine Bewertung des erwarteten Informationsgewinns vorzunehmen, muss eine Gütefunktion $R(\cdot)$ definiert werden. Dafür werden wir im Folgenden zwei informationstheoretische Gütefunktionen betrachten, die in der Sensoreinsatzplanung Verwendung finden und ihre jeweiligen Einschränkungen aufzeigen.

3.1 Entropie

Für diskrete Zufallsvariablen mit Bildbereich \mathcal{H} ist die Shannon–Entropie [3, 4] definiert als

$$H(\underline{\mathbf{X}}) := - \sum_{x \in \mathcal{H}} p(x) \log p(x) .$$

Für sie gilt unter anderem, dass ihr Wert nach oben durch die Entropie der Gleichverteilung auf \mathcal{H} beschränkt ist. Es gilt also

$$H(\underline{\mathbf{X}}) \leq \log |\mathcal{H}| ,$$

wobei $|\mathcal{H}|$ die Anzahl der Elemente in \mathcal{H} ist. Die Gleichheit gilt hier genau dann, wenn $\underline{\mathbf{X}}$ gleichverteilt auf \mathcal{H} ist [4]. Eine weitere Eigenschaft der

diskreten Entropie ist ihre Positivität, die direkt aus der Tatsache folgt, dass die Wahrscheinlichkeiten $p(x)$ immer in dem Intervall $[0, 1]$ liegen und deshalb $p(x) \log p(x) \leq 0$ ist.

Demgegenüber ist die Entropie einer kontinuierlichen Zufallsvariablen $\underline{\mathbf{X}} \sim f$ mit kontinuierlicher Wahrscheinlichkeitsdichte f definiert als

$$H(\underline{\mathbf{X}}) := E_{\underline{\mathbf{X}}} \{-\log f(\underline{x})\} = - \int f(\underline{x}) \log f(\underline{x}) d\underline{x} .$$

Diese Erweiterung der Entropie auf kontinuierliche Wahrscheinlichkeitsdichten ist keine direkte Verallgemeinerung der Shannon-Entropie [4] und besitzt einige ihrer grundlegenden Eigenschaften nicht. So kann die kontinuierliche Entropie zum Beispiel negative Werte annehmen. Ist $\underline{\mathbf{X}}$ gleichverteilt auf $[0, \frac{1}{n}]$, so erhält man für die kontinuierliche Entropie

$$H(\underline{\mathbf{X}}) = - \int_0^{\frac{1}{n}} n \cdot \log n \, dx = - \log n ,$$

was für $n > 1$ negativ ist. Neben der Shannon-Entropie und der kontinuierlichen Entropie finden auch aus ihnen abgeleitete Größen, wie die Kullback-Leibler-Divergenz [4, 5] oder die Transinformation [4], Anwendung in der Sensoreinsatzplanung [6–11]. Die Zielfunktion $R(\cdot)$ in der Sensoreinsatzplanung ist dann die erwartete Entropie der posterioren Dichte nach Messungen von Sensoren \underline{s}_k . Diese ist dann über die möglichen Messkonfigurationen zu optimieren.

3.2 Fisher-Information

Eine weitere informationstheoretische Gütefunktion, die zum Beispiel Anwendung in der schritthaltenden Lokalisierung findet [12, 13], ist die Fisher-Information. Sie ist definiert auf Räumen von Wahrscheinlichkeitsdichten $f_{\underline{\xi}}$

$$\mathcal{B} = \{f_{\underline{\xi}} = f(\underline{x}, \underline{\xi}) | \underline{\xi} \in \Lambda \subset \mathbb{R}^n\}$$

mit endlicher Parametrisierung $\underline{\xi} \in \Lambda$. Dabei muß gefordert werden, dass die Abbildung

$$\underline{\xi} \rightarrow f_{\underline{\xi}}$$

injektiv ist und

$$\underline{\xi} \rightarrow f(\underline{x}, \underline{\xi})$$

für jedes \underline{x} unendlich oft differenzierbar ist [14]. Ein Beispiel für eine Parametrisierung sind Gaußdichten, die über ihren Mittelwert und die Kovarianzmatrix charakterisiert werden. Für eine Dichte $f(\underline{x}, \underline{\xi}) \in \mathcal{B}$ ist dann ein Eintrag $g_{ij}(\underline{\xi})$ in der Fisher-Informationsmatrix definiert als

$$g_{ij}(\underline{\xi}) := \int \partial_{\xi_i} \log f(\underline{x}, \underline{\xi}) \cdot \partial_{\xi_j} \log f(\underline{x}, \underline{\xi}) \cdot f(\underline{x}, \underline{\xi}) d\underline{x} .$$

Die Fisher-Information ist ein Maß für die Information, die eine Zufallsvariable $\underline{X} \sim f(\underline{x}, \underline{\xi})$ über den Parameter $\underline{\xi}$ enthält. Der Nachteil bei ihrer Verwendung ist, dass eine endliche Parametrisierung der betrachteten Wahrscheinlichkeitsdichten vorliegen und dass diese Parametrisierung abgeschlossen unter den Operationen der Zustandsschätzung aus Kapitel 2.2 sein muß. Dies ist für nichtlineare Systeme nicht gegeben, da im nichtlinearen Fall nach der Prädiktion oder Filterung beispielsweise einer Gaußdichte nicht mehr gewährleistet ist, dass das Ergebnis wieder eine Gaußdichte darstellt.

4 Log-Ratio-Informationsmaß

In diesem Kapitel wird das Log-Ratio-Informationsmaß für kontinuierliche Wahrscheinlichkeitsdichten vorgestellt. Es hat den Vorteil, dass keine endliche Parametrisierung der unterlagerten Dichten erforderlich ist und es eine Norm auf einer bestimmten Klasse von Wahrscheinlichkeitsdichten darstellt. So sind Eigenschaften wie zum Beispiel Positivität intrinsisch.

4.1 Motivation

In diesem Ansatz werden Dichten als Elemente eines normierten Vektorraums \mathcal{A}^2 aufgefasst [15] und seine Norm als skalares Maß für den Informationsgehalt einer Dichte interpretiert.

In einem allgemeinen Vektorraum bestimmt die Norm die Länge eines Vektors, also wie weit dieser Vektor von dem eindeutig bestimmten Nullvektor des Raums entfernt ist. Der Nullvektor in \mathcal{A}^2 ist die Dichte der



Abbildung 11.1: Die Informationsfusion von zwei Sensoren (linke Seite) nach der Bayes Formel kann als Addition in einem Vektorraum aufgefasst werden (rechte Seite). Die Log-Ratio-Information der fusionierten Dichte $f \oplus g$ ist niemals größer als die Log-Ratio-Information der Teildichten f und g .

Gleichverteilung, die diejenige Verteilung ist, die am wenigsten Information über den Ausgang eines Zufallsexperiments bietet, da jeder Ausgang des Experiments gleich wahrscheinlich ist. Wenn wir also die Längen von Wahrscheinlichkeitsdichten in \mathcal{A}^2 messen, bestimmen wir, wie weit eine Wahrscheinlichkeitsdichte von der Dichte der Gleichverteilung entfernt ist.

Da das Log-Ratio-Informationsmaß eine Norm ist, wird außerdem gewährleistet, dass es für alle Dichten positive Werte annimmt. Zudem ist eine Form der Dreiecksungleichung unter der Bayesschen Datenfusion gegeben (s. Abb. 11.1). Diese Dreiecksungleichung ermöglicht es, obere Schranken für den Informationsgehalt einer fusionierten Dichte anzugeben.

4.2 Definition und Eigenschaften

In diesem Kapitel werden wir das Log-Ratio-Informationsmaß \mathbf{N} definieren und seine wichtigsten Eigenschaften nennen.

Definition 1 Sei $f : \Omega \rightarrow \mathbb{R}$ eine nicht-negative Funktion und $\log f \in L^2(\Omega)$. Die Log-Ratio-Information von f ist durch

$$\mathbf{N}(f) := \sqrt{\int_{\Omega} \int_{\Omega} \left[\log \left(\frac{f(\underline{x})}{f(\underline{y})} \right) \right]^2 d\underline{x} d\underline{y}} \quad (11.4)$$

gegeben.

In der Definition wird gefordert, dass der Logarithmus der Dichte f quadratintegrierbar über Ω sein muß, um sicher zu gehen, dass alle Inte-

grale existieren. Ein ähnliches Vorgehen wäre zu fordern, dass die Integrale in (11.4) wie in der Definition der kontinuierlichen Entropie existieren [4]. Die wesentlichen Eigenschaften des Log-Ratio-Informationsmaßes werden durch folgendes Theorem beschrieben.

Theorem 1 *Das Informationsmaß aus Definition 1 erfüllt die folgenden Eigenschaften für alle Dichten f und g wie in Definition 1*

- $N(f) \geq 0$,
- $N(f) = 0$ genau dann, wenn f die Dichte der Gleichverteilung auf Ω ist und
- $N(f \cdot g) \leq N(f) + N(g)$.

Der Beweis folgt aus der Konstruktion des Vektorraums und findet sich in [15].

Obwohl das Log-Ratio-Informationsmaß durch ein Doppelintegral über den Raum Ω definiert ist, gibt es eine Reformulierung mit einfachen Integralen, die die Berechenbarkeit vereinfacht.

Theorem 2 *Sei f eine Funktion gemäß Definition 1 und Ω beschränkt, dann gilt für $N(f)$*

$$N(f)^2 = 2\mu(\Omega) \int_{\Omega} [\log f(\underline{x})]^2 d\underline{x} - 2 \left[\int_{\Omega} \log f(\underline{x}) d\underline{x} \right]^2 ,$$

wobei μ das Maß auf Ω ist.

Wie sich das Log-Ratio-Informationsmaß verhält, zeigen folgende Beispielrechnungen:

1. Sei $\mathbf{X}_1 \sim f(x)$ eine exponentiell verteilte Zufallsvariable mit Dichte

$$f(x) = \lambda e^{-\lambda x} ,$$

definiert auf der positiven reellen Halbachse $[0, \infty)$. Der Logarithmus von f ist offensichtlich in $L^2([0, a])$ für jedes $a < \infty$. Für ein festes a ist

$$N(f)^2 = \lambda^2 \int_0^a \int_0^a (y - x)^2 dx dy .$$

Eine Maximierung von $N(f)^2$ führt also zu einer Minimierung der Varianz $\frac{1}{\lambda^2}$.

2. Ist $\mathbf{X}_2 \sim G(x)$ eine Zufallsvariable mit Gaußdichte

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{\sigma^2}\right),$$

so gilt für jedes beschränkte $\Omega = [a, b] \subset \mathbb{R}$, dass der Logarithmus von G quadratintegrierbar ist,

$$\log G = -\frac{x^2}{\sigma^2} - \log \frac{1}{\sqrt{2\pi}\sigma}$$

ein Polynom zweiten Grades ist. Die Log-Ratio-Information von G ist

$$\mathbf{N}(G)^2 = \frac{1}{\sigma^2} \int_a^b \int_a^b (y^2 - x^2) dx dy,$$

und für festes Intervall $[a, b]$ entspricht eine Maximierung von $\mathbf{N}(G)$ einer Minimierung der Standardabweichung $\frac{1}{\sigma}$.

5 Sensoreinsatzplanung

In diesem Kapitel werden wir das Log-Ratio-Informationsmaß auf die Sensoreinsatzplanung anwenden, indem wir es als eine informationstheoretische Gütefunktion auffassen. Aus Gründen der Übersichtlichkeit betrachten wir nur den Fall, dass zu jedem Zeitschritt k ein Sensor s_k für Messungen ausgewählt werden soll. Der Ansatz kann aber ohne Weiteres auf den Fall einer gesamten Sensorkonfiguration $\underline{s}_k = [s_{k1}, \dots, s_{kn}]$ verallgemeinert werden.

5.1 Auswahl von Sensoren

Die Aufgabe des Sensormanagers ist es, Sensoren für Messungen auszuwählen, deren Messungen einen hohen Informationsgehalt über den Zustand des Systems haben. Wir bezeichnen mit f_k die Dichte, die die gegenwärtige Zustandsschätzung beschreibt, und mit f_k^p die Dichte, die wir erhalten, nachdem wir die gegenwärtige Schätzung mit dem Prädiktionsschritt (11.2) fortschreiben. Die Dichte $f_{s_k}^e$ bezeichne die posteriore Dichte

$$f_{s_k}^e(\underline{x}_k | \hat{\underline{z}}_{s_k}) := \frac{1}{c_k} f_k^p(\underline{x}_k) \cdot f_{s_k}^L(\hat{\underline{z}}_{s_k} | \underline{x}_k),$$

die erhalten wird, nachdem die priore Dichte f_k^p mit der Likelihood $f_{s_k}^L$ des Sensors s_k gefiltert wurde.

Den erwarteten Informationsgewinn nach Filterung der aktuellen Zustandsschätzung mit Messungen z_{s_k} von Sensor s_k wird als

$$E_{z_{s_k}} \{ \mathbf{N}(f_{s_k}^e(\cdot | z_{s_k})) \} \quad (11.5)$$

definiert. Die Schreibweise $\mathbf{N}(f_{s_k}^e(\cdot | z_{s_k}))$ bedeutet, dass die Log-Ratio-Information in Abhängigkeit von z_{s_k} ausgewertet wird. Eine Maximierung von (11.5) führt zu einer Maximierung des erwarteten Informationsgewinnes im nächsten Zeitschritt. Bei einem modell-prädiktiven Ansatz zur Planung über einen gewissen Horizont wird diese Gütefunktion in die rekursive Gleichung zur Bestimmung einer optimalen Lösung eingefügt [2]. Die Lösung nach dem Bellmanschen Optimalitätsprinzip für den maximalen Informationsgewinn über N Zeitschritte berechnet sich rekursiv aus

$$V_k(f_k) := \max_{s_k} [E_{z_{s_k}} \{ \mathbf{N}(f_{s_k}^e(\cdot | z_{s_k})) + V_{k+1}(f_{s_k}^e(\cdot | z_{s_k})) \}] \quad (11.6)$$

mit terminalem Gewinn

$$V_N(f_k) := \max_{s_N} [E_{z_{s_N}} \{ \mathbf{N}(f_{s_N}^e(\cdot | z_{s_N})) \}] \quad (11.7)$$

im letzten Zeitschritt. Durch die vorausschauende Planung fließt das zukünftige Verhalten des Phänomens unter Berücksichtigung des Systemrauschens bereits bei der Auswahl der Sensoren ein. Das folgende Beispiel zeigt, dass durch diesen Ansatz eine hohe Schätzqualität unter Verwendung nur eines Sensors pro Zeitschritt erreicht werden kann, was zu einer erheblichen Erhöhung der Lebensdauer des Netzwerks führt.

5.2 Beispiel

Wir betrachten ein Beispiel bestehend aus einem Sensornetzwerk mit drei Winkelsensoren und zwei Abstandssensoren [11]. Die Aufgabe des Sormanagers in diesem Beispiel ist es, in jedem Zeitschritt einen Sensor auszuwählen, um ein Fahrzeug über zehn Zeitschritte zu lokalisieren. Die quadratische Abweichung der geschätzten zu der wahren Position des Fahrzeugs wird in Abb. 11.2 als durchgezogene Linie dargestellt. Diese Abweichung wird verglichen mit der Qualität der Zustands-

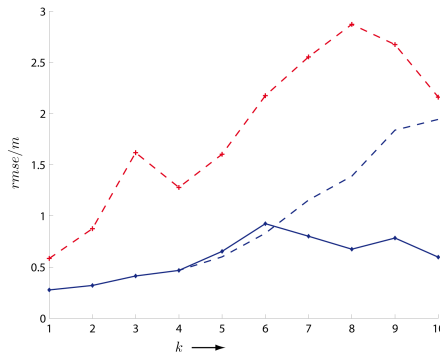


Abbildung 11.2: Die durchschnittliche quadratische Abweichung (rmse) der Zustandsschätzung über 20 Testläufe mit einem in jedem Zeitschritt zufällig ausgewählten Sensor (rote gestrichelte Linie mit Kreuzen), der Entropie (blaue gestrichelte Linie) und der Log-Ratio-Information als Gütefunktion (blaue durchgezogene Linie).

schätzung, die ein Manager erreicht, der zufällig einen Sensor zum Messen auswählt (gestrichelte Linie). Es ist dabei deutlich zu sehen, dass die Zustandsschätzung für die Log-Ratio-Information wesentlich genauer ist.

6 Ausblick

Das vorgestellte Log-Ratio-Informationmaß weist Eigenschaften wie Positivität und die Erfüllung der Dreiecksungleichung auf. Diese erlauben es, sowohl obere als auch untere Schranken für den optimalen Informationsgewinn anzugeben. Dadurch wird eine ganze Klasse von neuartigen Verfahren zur Bestimmung der optimalen Messkonfigurationssequenz erschlossen und eine Bewertung der so gefundenen Lösungen ermöglicht.

Literatur

1. D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2007.
2. R. E. Bellman, „On the Theory of Dynamic Programming“, *Proceedings of the National Academy of Sciences*, Vol. 38, S. 716–719, 1952.

3. C. E. Shannon, *The mathematical theory of communication*. Urbana: Univ. of Ill. Press, 1949.
4. T. M. Cover und J. A. Thomas, *Elements of Information Theory*, Ser. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.
5. S. Kullback und R. A. Leibler, „On Information and Sufficiency“, *Annals of Mathematical Statistics*, Vol. 22, Nr. 2, S. 79–86, 1951.
6. A. O. Hero, C. M. Kreucher und D. Blatt, *Foundations and Applications of Sensor Management*, Ser. Signals and Communication Technology. Springer, 2008, Kap. Information Theoretic Approaches to Sensor Management, S. 33–59.
7. E. Ertin, J. W. Fisher und L. C. Potter, „Maximum Mutual Information Principle for Dynamic Sensor Query Problems“, in *Information Processing in Sensor Networks (IPSN)*, 2003, S. 405–416.
8. A. S. Chhetri, D. Morrell und A. Papandreou-Suppappola, „Nonmyopic Sensor Scheduling and its Efficient Implementation for Target Tracking Applications“, *EURASIP Journal on Applied Signal Processing*, Vol. 2006, S. 1–18, 2006.
9. J. L. Williams, „Information Theoretic Sensor Management“, Dissertation, Massachusetts Institute of Technology, Feb. 2007.
10. F. Zhao, J. Shin und J. Reich, „Information-driven Dynamic Sensor Collaboration for Tracking Applications“, *IEEE Signal Processing Magazine*, Vol. 19, Nr. 2, S. 61–72, 2002.
11. M. F. Huber, „Probabilistic Framework for Sensor Management“, Dissertation, Universität Karlsruhe(TH), 2009.
12. M. L. Hernandez, T. Kirubarajan und Y. Bar-Shalom, „Multisensor Resource Deployment Using Posterior Cramér-Rao Bounds“, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 40, S. 399–416, 2004.
13. M. L. Hernandez, „Optimal Sensor Trajectories in Bearings-Only Tracking“, in *Proceedings of the 7th International Conference on Information Fusion (Fusion 2004)*, 2004.
14. S. Amari und H. Nagaoka, *Methods of Information Geometry*, Ser. Translation of Mathematical Monographs. American Mathematical Society, 2000, Vol. 191.
15. J. J. Egozcue, J. L. Diaz-Barrero und V. Pawlosky-Glahn, „Hilbert Space of Probability Density Functions Based on Aitchison Geometry“, *Acta Mathematica Sinica, English Series*, Vol. 22, Nr. 4, S. 1175–1182, July 2006.