

Entropie in der Informationstheorie

Entropy in Information Theory

Bernhard C. Geiger, Gernot Kubin, Christoph Temmel, Wolfgang Woess

Wissen Sie, wie viel Bit Ihre MP3-Musikdateien an Speicherplatz auf Ihrem Smartphone benötigen? Und erinnern Sie sich noch, welche Unsummen die Mobilfunkbetreiberinnen und -betreiber für das benötigte Funkspektrum bezahlt haben, um zunächst Sprachtelefonie und später eine Vielzahl digitaler Kommunikationsdienste mit immer höheren Datenraten anzubieten? Dann haben Sie schon mit der Entropie, dem Maß des Informationsgehalts von Daten, Prozessen und Signalen, zu tun gehabt. Mathematik und Signalverarbeitung arbeiten zusammen, um das Fundament von ICC (Information, Communication & Computing) in Bezug auf Theorie und Anwendung zu erweitern.

Was ist Entropie? Im Kern ist die Entropie ein Maß für den mittleren Informationsgehalt einer zufälligen Nachricht und damit für unser Unwissen über den Inhalt einer noch nicht gelesenen Nachricht. Wenn wir uns Letztere als ein Rätselwort vorstellen, so gibt es eine Minimalzahl an Ja/Nein-Fragen, die nötig sind, um das Rätsel zu lösen. Da aber das Rätselwort auch noch zufällig ist, also Wahrscheinlichkeiten unterliegt, müssen wir den gewichteten Mittelwert über die nötige Zahl an Ja/Nein-Fragen nehmen. Das führt zur Formel für die Entropie der Wahrscheinlichkeitsverteilung $H(p) = -\sum p(x) \log_2 p(x)$. Entropie spielt in der Thermodynamik eine ebenso große Rolle wie in der Informationstheorie; die Verbindung erklärt der Maxwellsche Dämon, dem wir aber hier nicht näher treten wollen.

Am Institut für Mathematische Strukturtheorie ist Entropie in verschiedenen Formen Teil von theoretischen Untersuchungen von Zufallsprozessen. Die am Ende gelisteten Arbeiten von W. Huss, E. Sava und W. Woess¹, L. Gilch² sowie V. A. Kaimanovich und W. Woess³ seien hier als Beispiele genannt. In allen Fällen geht es um Entropie-Raten, typischerweise von Markovketten, das sind Zufallsprozesse mit endlich (oder abzählbar) vielen möglichen

Do you know how many bits of data memory your mp3 music files occupy on your smartphone? And do you still remember the crazy amounts paid by the mobile operators for the radio spectrum required for offering, first, voice telephony, and then later, a variety of digital communications services with ever increasing data rates? Then you have already encountered entropy as the measure of information content of data, processes, and signals. Mathematics and signal processing team up to extend the theory and application of this foundation of ICC (Information, Communication & Computing).

What is entropy? In a nutshell entropy is a measure of the average information content of a random message and, thereby, of our ignorance about the content of a yet unread message. If we imagine the latter as a mystery word then there is a minimal number of yes-no questions needed to unravel the mystery. But as the mystery word itself is random, that is, governed by probabilities, we need to take the weighted average over the necessary number of yes-no questions. This leads to the expression for the entropy of a probability distribution $H(p) = -\sum p(x) \log_2 p(x)$. Entropy plays an equally important role in thermodynamics as well as in information theory. This connection is explained by Maxwell's demon which we would rather not examine more closely here.

At the Institute for Mathematical Structure Theory, entropy in various forms makes up part of the theoretical investigations of random processes. For these, the articles listed at the end by W. Huss, E. Sava and W. Woess¹, L. Gilch², as well as by V. A. Kaimanovich and W. Woess³ may serve as examples. In all cases, it is a question of entropy rates, typically of "Markov chains" – random processes with a finite (or countable) set of possible states and a memory which extends only

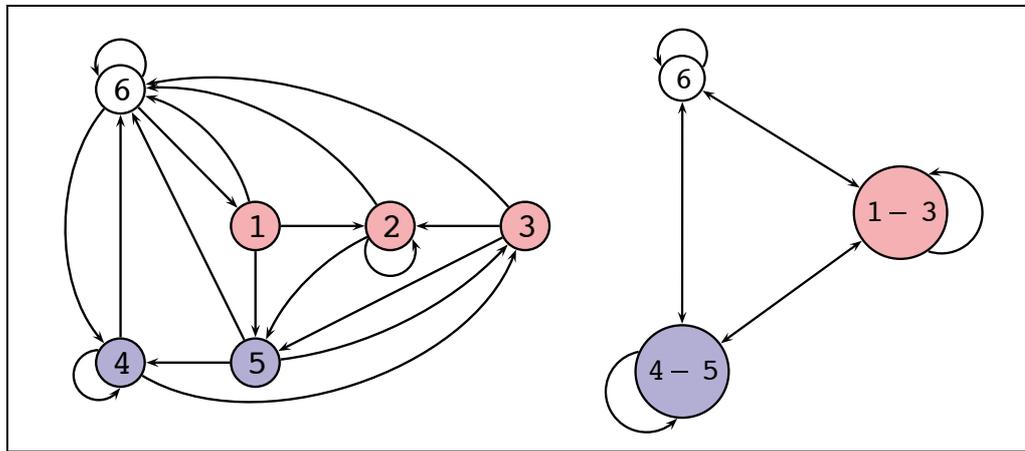


Bernhard C. Geiger ist Universitätsassistent am Institut für Signalverarbeitung und Sprachkommunikation. Sein Forschungsschwerpunkt liegt im Bereich der Informationstheorie; in seiner Dissertation untersucht er den Informationsverlust in deterministischen Systemen.

Bernhard C. Geiger is a research and teaching associate at the Signal Processing and Speech Communication Laboratory. His research interests include information theory, and his doctoral thesis investigates information loss in deterministic systems.

Abb. 1: Zustandszusammenfassung im Graphen einer Markovkette.

Fig. 1: A lumping of states in the graph of a Markov chain.



© TU Graz/Geiger



Gernot Kubin leitet die Institute für Signalverarbeitung und Sprachkommunikation sowie für Technische Informatik, koordiniert die Doctoral School Informations- und Kommunikationstechnik und leitet das FoE „Information, Communication & Computing“. Seine Forschungsinteressen liegen in der nichtlinearen Signalverarbeitung, Computational Intelligence sowie der Sprach- und Audiokommunikation.

Gernot Kubin is head of the Signal Processing & Speech Communication and the Technical Informatics Laboratories, coordinator of the Doctoral School Information and Communications Engineering, and head of the FoE “Information, Communication & Computing“. His research interests are in nonlinear signal processing, computational intelligence, speech and audio communication.

Zuständen und einem Gedächtnis, das nur einen Zeitschritt zurückreicht. Diese Prozesse spielen in der angewandten Informationstheorie die zentrale Rolle: Sie sind der Ausgangspunkt für die Kooperation zwischen den beiden Instituten.

Am Institut für Signalverarbeitung und Sprachkommunikation werden neue theoretische Ansätze zur Charakterisierung informationsverarbeitender Systeme mithilfe einer Entropiebilanz entwickelt, ähnlich wie mithilfe einer Energiebilanz elektrotechnische oder mechanische Systeme charakterisiert werden können. Auch bei der Optimierung von selbstlernenden Systemen werden informationstheoretische Kostenfunktionen eingesetzt. Letztlich bilden Markovketten zur Modellierung der Abfolge sprachlicher Zeichen (Laute, Silben, Wörter usw.) die Grundlage für die automatische Spracherkennung, wie sie zuletzt auch bei Smartphones populär geworden ist (vgl. z. B. „Siri“). Siehe dazu auch die unten gelisteten Arbeiten von B. Geiger und G. Kubin^{4,5} sowie P. Mowlaee⁶ et al.

Markovketten auf der Spur

Im Rahmen einer Kooperation der beiden Institute wird nun an folgendem informationstheoretischen Problem in der Theorie der Markovketten gearbeitet (siehe B. Geiger und C. Temmel^{7,8}):

Wenn man eine Markovkette dadurch vereinfacht, dass man mehrere Zustände zusammenfasst, bekommt man einen Prozess mit weniger möglichen Zuständen [HMM]¹. Im Allgemeinen verliert man dadurch aber nicht nur die Markov-Eigenschaft, sondern auch einen beträchtlichen Teil der Information: Die Entropie-Rate wird verringert. Die zentrale Frage der Kooperation zwischen dem Institut für Mathematische Strukturtheorie und dem Institut für Signalverarbeitung und Sprachkommunikation war die Charakterisierung jener Zusammenfassungen, welche die Entropie-Rate erhalten. Als Hauptresultat bewiesen wir, dass der Erhalt der Entropie-Rate gleichbedeutend der viel stärkeren Eigenschaft der

a single time step into the past. These processes play the key role in applied information theory, and are the starting point for the cooperation between the two institutes.

At the Signal Processing and Speech Communication Laboratory, new theoretical approaches to the characterization of information processing systems using an entropy balance are being developed, similar to the characterization of electrical or mechanical systems using an energy balance. Information theoretic cost functions are employed for the optimization of machine learning systems, too. Finally, Markov chains modeling sequences of spoken language symbols (speech sounds, syllables, words etc.) form the basis for automatic speech recognition as recently popularized on smartphones (cf. e.g., “Siri”). Please refer to the articles listed below by B. Geiger and G. Kubin^{4,5} as well as P. Mowlaee⁶ et al.

Hot on Markov chains

In the framework of a cooperation between the two institutes, the following information theoretic problem in the theory of Markov chains is addressed (cf. B. Geiger und C. Temmel^{7,8}):

If we simplify a Markov chain by lumping together multiple states, we obtain a process with fewer possible states [HMM]¹. In general, we do not only lose the Markov property from this, but also a sizeable part of the information: the entropy rate is reduced. The key question in the cooperation between the Institute for Mathematical Structure Theory and the Signal Processing and Speech Communication Laboratory concerned the characterization of these lumpings which conserve entropy rate. Our proven main result is that the conservation of the entropy rate is equivalent to the much stronger property of invertibility. In other words, no information is lost *on average* by lumping together multiple states if and only if the original Markov chain can be reconstructed from

¹ [HMM] ist auch als Hidden Markov Model bekannt.

¹ [HMM] also known as Hidden Markov Model.

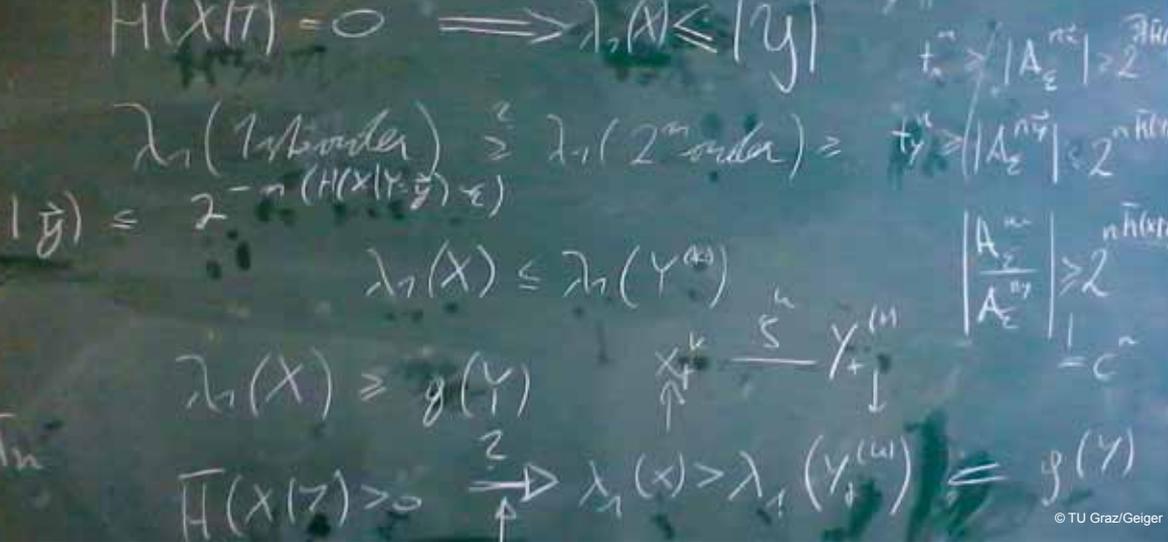


Abb. 2: Markovketten auf der Spur.
Fig. 2: Hot on Markov chains.



Christoph Temmel war wissenschaftlicher Mitarbeiter am Institut für Mathematische Strukturtheorie und ist nun Postdoc an der VU Amsterdam. Er forscht an der Schnittstelle zwischen Kombinatorik, diskreter Stochastik und statistischer Mechanik.

Christoph Temmel used to be a scientific assistant at the Institute for Mathematical Structure Theory and is now a postdoc at the VU Amsterdam. He carries out research at the frontier between combinatorics, discrete stochastics and statistical mechanics.

Umkehrbarkeit ist. Anders ausgedrückt: Durch eine Zusammenfassung geht genau dann *im Mittel* keine Information verloren, wenn man vom Prozess mit wenigen Zuständen die ursprüngliche Markovkette rekonstruieren kann. Nebenbei formulierten wir noch hinreichende Bedingungen dafür, dass die Zusammenfassung nicht nur die Information erhält, sondern auch noch die Markov-Eigenschaft einer bestimmten Ordnung besitzt. Dies ermöglicht die Aggregation komplexer Markov-Modelle auf ein kleineres Modell, welches immer noch einfach zu simulieren ist.

Thema des Beitrages ist eine kürzlich entstandene Kooperation auf theoretischem Gebiet zwischen Instituten zweier Fakultäten: ein zartes Pflänzchen, das ohne spektakuläres Auftreten die Verbindung verschiedener Fachbereiche im Field of Expertise „Information, Communication & Computing“ der TU Graz aufzeigt. ■

the process with fewer states. As a corollary, we formulate sufficient conditions for this lumping together to not only conserve information but also to still keep the Markov property of a certain order. This facilitates the aggregation of complex Markov models into smaller models which continue to be easily simulated.

This article came about due to a recent cooperation in a theoretical domain between institutes of two faculties. This demonstrates the productive and appreciative approach shared by different scientific areas in the Graz University of Technology Field of Expertise “Information, Communication & Computing”. ■



Wolfgang Woess ist Leiter des Instituts für Mathematische Strukturtheorie (Math C) an der TU Graz. Seine Forschungsinteressen umfassen Zufallsprozesse auf Graphen und Gruppen, die Struktur von unendlichen Graphen und Gruppen, formale Sprachen, Theorie der Ränder und harmonische Funktionen sowie Potenzialtheorie.

Wolfgang Woess is head of the Institute of Mathematical Structure Theory (Math C). His research interests are random processes on graphs and groups, structure of infinite graphs, groups and formal languages, boundary theory and harmonic functions and potential theory.

Literatur/References:

¹ Wilfried Huss, Ecaterina Sava and Wolfgang Woess: Entropy sensitivity of languages defined by infinite automata, via Markov chains with forbidden transitions, *Theoretical Computer Science* 411 (2010), 3917–3922.

² Lorenz Gilch: Asymptotic Entropy of Random Walks on Free Products. *Electronic Journal of Probability* 16 (2011), 76–105.

³ V. A. Kaimanovich and W. Woess: Boundary and entropy of space homogeneous Markov chains, *Ann. Probab.* 30 (2002) 323–363.

⁴ Bernhard C. Geiger and Gernot Kubin: Information Measures for Deterministic Input-Output Systems. Preprint TUGraz (2013), arXiv:1303.6409 [cs.IT].

⁵ Bernhard C. Geiger and Gernot Kubin: Signal Enhancement as Minimization of Relevant Information Loss. *ITG Conf. on Systems, Communication and Coding (Munich, 2012)* arXiv:1205.6935 [cs.IT].

⁶ P. Mowlaee et al., The 2nd CHIME Speech Separation and Recognition Challenge: Approaches on Single-Channel Speech Separation and Model-Driven Speech Enhancement, in *Proc. 2nd CHIME Speech Separation and Recognition Challenge*, (Vancouver, 2013).

⁷ Bernhard C. Geiger, Christoph Temmel: Lumpings of Markov chains and entropy rate loss. Preprint, TU Graz (2012). arXiv:1212.4375 [cs.IT].

⁸ Bernhard C. Geiger, Christoph Temmel: Information-preserving Markov aggregation. Preprint, TU Graz (2013). arXiv:1304.0920 [cs.IT].