



Modeling Human Frailty with a Smart Home-Based Approximation of Entropy

Katherine A. Wuestney
Washington State University
Vancouver, Washington, USA
katherine.graham@wsu.edu

Diane J. Cook
Washington State University
Pullman, Washington, USA
djcook@wsu.edu

Beiyu Lin
University of Nevada Las Vegas
Las Vegas, USA
beiyu.lin@unlv.edu

Roschelle L. Fritz
Washington State University
Vancouver, Washington, USA
shelly.fritz@wsu.edu

ABSTRACT

The global population is aging. These older adults face changes in health that result in increasing frailty. Digital biomarkers created by passive, continuous sensors offer an early indicator of impending frailty that can be used to delay or reverse frailty. Building on the notion of a human as a complex system, we introduce and compare three methods to model and estimate the complexity of indoor human behavior. Each method offers potential benefits for estimating human frailty. We introduce a formalization of the approaches, extend their use for arbitrary-size sensor suites, and demonstrate how they can be used to visualize and calculate a person's behavioral complexity based on smart home data collected continuously for an older adult subject.

CCS CONCEPTS

• **General and reference** → **Experimentation**; • **Mathematics of computing** → **Information theory**; *Markov processes*; **Time series analysis**; • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

behavior complexity, data complexity, frailty, entropy, smart homes, CGR/USM

ACM Reference Format:

Katherine A. Wuestney, Beiyu Lin, Diane J. Cook, and Roschelle L. Fritz. 2023. Modeling Human Frailty with a Smart Home-Based Approximation of Entropy. In *14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '23)*, September 3–6, 2023, Houston, TX, USA. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3584371.3613017>

1 INTRODUCTION

Aging involves an array of biopsychosocial changes which interact in multidimensional, dynamic pathways to produce a range of

common symptom profiles, or syndromes, frequently experienced by older adults. There is a growing consensus among researchers that the heterogeneous trajectories of aging syndromes cannot be fully understood without a complex systems approach [5]. Such an approach views humans as complex systems with multiple levels of factors interacting in nonlinear ways. These complex dynamics result in the emergence of system properties, such as homeostasis or resilience, which are greater than the sum of the lower-scale factors. From a complex systems framework, loss of complexity in any part of the human system may lead to the same emergent outcomes of frailty and decline [12].

Mounting evidence suggests that the complexity of a person's observable behavior signals (e.g. gait variability and standing postural sway among others) could serve as a 'digital biomarker' of the body's level of frailty [5]. Digital biomarkers created passively via ubiquitous sensing technologies offer new opportunities for continuous, minimally obtrusive, and remotely delivered assessment of health conditions in ecologically valid real-world settings [4]. Frailty, a geriatric syndrome characterized by functional decline across multiple physiologic systems and reduced capacity to overcome internal and external stressors, can be conceptualized as a complex system on the brink of failure [15]. Measuring change of complexity of the human system's signals would then be an indicator of how much complexity has been lost and, therefore, how close the system is to failure (i.e., frailty). This Loss of Complexity Hypothesis may explain why an individual with health deficits related to cardiovascular disease develops the same phenotypic frailty presentation as another individual with a diagnosis of COPD and no cardiovascular disease or another individual with mild deficits in a number of physiologic domains but no formal chronic disease diagnosis [8, 12].

The progression of frailty severity can be slowed if not altogether reversed, making early identification and intervention of worsening frailty a critical function of geriatric healthcare [9]. Compared to conventional health measures of frailty, which are not amenable to continuous measurement and can be time consuming, expensive, and physically invasive, digital biomarkers based on purely passively-collected smart home sensor data would further minimize patient burden and may extend the reach of community prevention and early intervention of frailty.

In this paper, we propose measuring the complexity of human indoor activity using smart home sensor data, defined as the level



This work is licensed under a Creative Commons Attribution International 4.0 License.

BCB '23, September 3–6, 2023, Houston, TX, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0126-9/23/09.

<https://doi.org/10.1145/3584371.3613017>

of randomness and irregularity of sensor transition patterns, as a potential digital biomarker for detecting change in a person's frailty. The central hypothesis motivating this work is that the complexity of smart home movement trajectories of older adults diminishes with health decline and frailty. We introduce a novel approach to quantify and visualize the complexity of smart home data using universal sequence maps (USM) with the Rényi quadratic entropy measure (USM-Rényi) [18]. A USM is a bijective projection of a sequence from a discrete state space (e.g., sequence of sensor readings) to a real-valued state space. We expand the original formula for USM-Rényi to symbolic sequences of arbitrary alphabet size and compare its sensitivity to changes with two other entropy measures designed for short time series, Approximate Entropy and Sample Entropy, using Monte Carlo Simulations.

Our contributions include formalizations of behavior complexity measures and a new proof of USM-Rényi for arbitrary alphabet sizes. We demonstrate which entropy measure converges fastest and provides the narrowest error on short sample lengths. Additionally, we present a novel application of the USM algorithm to the visualization of smart home motion motifs at multiple scales. Our analysis demonstrates how these visualizations can easily discriminate days with and without novel complex activity (due to visitors and housekeeping) in a case example.

2 RELATED WORK

To date, few studies of digital biomarker technologies targeting frailty utilize an ambient sensor platform or passive monitoring, and none utilize a complex systems framework of aging [17]. Schütz et al. [2022] include entropy as one of 1,269 digital biomarkers they explored from smart home data for 45 older adults using a "holistic systems" machine learning approach to predict frailty among other outcomes. Of their markers, they found fridge entropy to be moderately negatively correlated ($r=-0.25$) with frailty. Because their entropy measure was based on the uncorrelated frequency of hourly readings, the approach misses higher-order behavior complexities such as transitions between locations.

Other smart home studies explore correlated entropy measures that do reflect these multi-order transitions. Unlike our study, these related analyses focus primarily on movement predictability, rather than health outcomes [19, 20]. Howedi et al. [2020] leveraged Sample Entropy and Approximate Entropy, two measures we analyze in this paper also, to identify periods of multi-occupancy in smart home data. In order to detect change in complexity of motion trajectories, we need measures which reflect multi-order complexity of motion trajectories that are reliable with relatively short time series. Of these studies, only Wang et al. [2023] explored the sensitivity of the measures to sequence length and none explored the sensitivity of the entropy measures to number of sensors (i.e., alphabet size), even though information entropy and its derivatives are known to be sensitive to both alphabet size and sequence length [7].

3 SMART HOME DATA

For our study, we analyze data collected by ambient sensors as part of the CASAS Smart Home in a Box [6]. As shown in Figure 1, sensors include motion detectors, magnetic door sensors, ambient

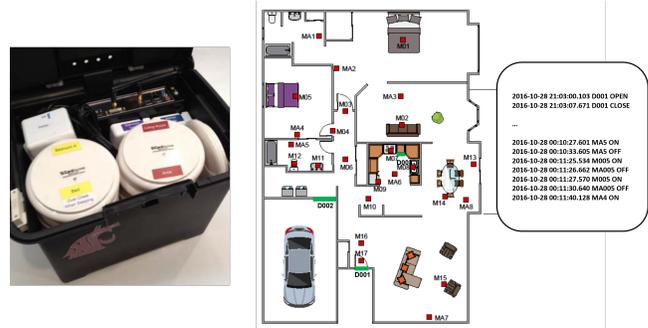


Figure 1: CASAS SHiB. Infrared motion detectors are installed inside the refrigerator and in each functional area of the house. Magnetic door closure sensors are attached to external doors and selected cabinets. Sensor readings are stored by the smart home for analysis.

light and temperature sensors. All sensors generate a reading whenever they detect a change in state (e.g., when motion starts or stops in a sensor's field of view.) Each sensor reading includes a corresponding date, time, sensor identifier, and sensed state (ON/OFF for motion, OPEN/CLOSED for doors, numeric values for light and temperature sensors). As in [19, 20] we construct our motion trajectories using the ON and OPEN readings, excluding OFF and CLOSE.

4 METHODOLOGY

We model human behavior based on movement trajectories captured by smart home sensors. We assume that a person's trajectory is a non-stationary stochastic process whose complexity changes over time in association with health status. We then can use the model to estimate the complexity of the underlying "human system" and the model is used as a basis for estimating a person's frailty. Throughout the paper, we represent a smart home sensor reading trajectory as a symbolic sequence, X , whose alphabet $S = \{s_1, \dots, s_d\}$ represents the set of sensors with cardinality d and x_i represents the reading generated at index i in sequence X . In our methods, we implement and compare three methods for estimating complexity of smart home sensor sequences: USM, Approximate Entropy, and Sample Entropy. We compare the performance of each method on short symbolic sequences using Monte Carlo simulation and demonstrate the potential of USM methodology in a smart home case study.

4.1 Method 1: Universal Sequence Maps (USM)

Universal sequence mapping maps a sequence with a discrete alphabet to a real-valued state space in \mathbb{R}^d . USMs function as generalized order-free Markov transition matrices of symbolic sequences (i.e., sequences of sensor readings). We compute an entropy estimate of a sequence X based on the density of USM coordinates.

Generating the USM space is based on the algorithm for chaos game representation (CGR, [11]). First, each symbol in a sequence is assigned to a vertex of a unit hypercube such that the vertex coordinates $\{Y\}$ of the USM hypercube are equal to the rows of

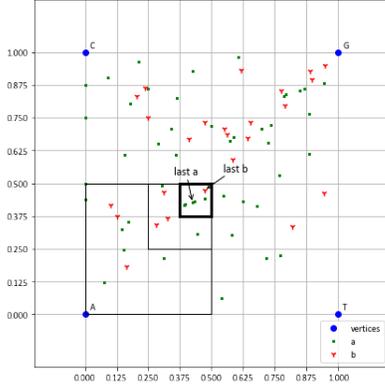


Figure 2: CGR plot of sequences with a shared suffix. The last 4 symbols of a are 'AGGA' and the last 4 of b are 'GGGA'. The highlighted sub-quadrant contains the coordinate of the sequences' last symbol, the quadrant size corresponds to the suffix length.

the $d \times d$ identity matrix, reflecting a one-hot encoding of S . Next, starting at the beginning of the sequence, the USM coordinate, a_i , for each symbol in X is computed iteratively as the midway point between the vertex associated with that symbol and the previous symbol's coordinate. This is given by Equation 1, where y_i is the vertex coordinate associated with the i^{th} symbol in X .

$$a_i = a_{i-1} + \frac{1}{2}(y_i - a_{i-1}), \quad i = 1, \dots, N \quad (1)$$

This approach ensures that the USM projection is always bijective, no matter the alphabet size [2]. Theoretically, this means that a single USM coordinate contains information of the entire sequence and conversely the entire sequence could be recovered from a single coordinate. USM is therefore a universal, non-parametric method for analyzing the statistical properties on multiple scales of symbolic sequences of any alphabet size. CGR/USM images, such as the one shown in Figure 2, reveal characteristic patterns related to the density of different k -grams in the sequence. This is because all coordinates within the same sub-quadrant will always share the same k preceding symbols. Therefore, the Euclidean distance between any two points in the USM does not indicate the nearness within the parent sequence but rather the length of a shared suffix. Because of this property, the USM map depicts all k -gram frequencies simultaneously. This makes the USM an *order free* generalization of a Markov Chain Transition Matrix, with the k^{th} order transition frequencies equal to the density of coordinates within the sub-quadrants with side lengths 2^{-k} [1].

Based on this property, Vinga and Almeida proposed a method for measuring the continuous Rényi entropy of 4-dimensional USM (i.e., to analyze sequences with an alphabet of size 4) based on calculating the density distribution of a USM with a Parzen kernel density estimate with Gaussian kernel density function [18]. The equation for this 4-dimensional USM-Rényi is given in Equation 2, where d_{ij} is the squared Euclidean distance between USM coordinates a_i and a_j and σ determines the size of the kernel.

$$H_2(USM) = -\ln \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{16\pi^2\sigma^4} \exp\left(-\frac{1}{4\sigma^2}d_{ij}\right) \quad (2)$$

We now offer a new proof that extends this Rényi equation for an alphabet of any size, d . Rényi entropy was introduced as a generalization of Shannon entropy and includes an order parameter α which determines the weighted contribution of improbable events to the overall entropy measure. The limit of Rényi entropy $\lim_{\alpha \rightarrow 1} H_\alpha(X)$ is Shannon's entropy measure. The formula for Rényi entropy of a continuous probability density function $f(x)$ is:

$$H_\alpha(X) = \frac{1}{1-\alpha} \ln \int f^\alpha(x) dx \quad (3)$$

We first substitute the kernel density equation of a spherical Gaussian kernel [18] for $f(x)$ in Equation 3:

$$\begin{aligned} H_2(USM) &= -\ln \int \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi)^{d/2} |\sigma^2 I_d|^{1/2}} e^{-\frac{1}{2}(x-a_i)^T (\sigma^2 I_d)^{-1} (x-a_i)} \right)^2 dx \\ &= -\ln \int \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{1}{(2\pi)^{d/2} |\sigma^2 I_d|^{1/2}} e^{-\frac{1}{2}(x-a_i)^T (\sigma^2 I_d)^{-1} (x-a_i)} \right) \\ &\quad \times \left(\frac{1}{(2\pi)^{d/2} |\sigma^2 I_d|^{1/2}} e^{-\frac{1}{2}(x-a_j)^T (\sigma^2 I_d)^{-1} (x-a_j)} \right) dx \end{aligned} \quad (4)$$

Applying the constant coefficient rule of integration and the summation rule of integration, we move the integral operator inside the summation. The entropy is now a convolution of the two Gaussians:

$$H_2(USM) = -\ln \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{(2\pi)^{d/2} |\sigma^2 I_d + \sigma^2 I_d|^{1/2}} \times e^{-\frac{1}{2}(a_i - a_j)^T (\sigma^2 I_d + \sigma^2 I_d)^{-1} (a_i - a_j)}$$

Because we are using a fixed kernel volume approach, σ^2 is held constant across all kernels, so we can further simplify $\sigma^2 I_d + \sigma^2 I_d$ to $2\sigma^2 I_d$. Then $|2\sigma^2 I_d|^{1/2} = (2^d \sigma^{2d})^{1/2} = 2^{d/2} \sigma^d$ and $(2\sigma^2 I_d)^{-1} = \frac{1}{2\sigma^2} I_d$. The former equality is due to the fact that the determinant of a $d \times d$ scalar matrix is the scalar constant raised to the d^{th} power, $|cI_d| = c^d$. The latter is due to the fact that the inverse of a diagonal matrix is a diagonal matrix whose principal diagonal is made up of the reciprocals of the elements of the original matrix, which in this case is $\frac{1}{2\sigma^2}$. Plugging these two simplified terms into the entropy equation, we can perform further simplifications based on the properties of matrix-vector multiplication involving scalars and the identity matrix:

$$\begin{aligned}
H_2(USM) &= -\ln \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{(2\pi)^{d/2} 2^{d/2} \sigma^d} e^{\left(-\frac{1}{2}(a_i - a_j)^T \frac{1}{2\sigma^2} I_d (a_i - a_j)\right)} \\
&= -\ln \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{(2\pi^{1/2} \sigma)^d} e^{\left(-\frac{1}{4\sigma^2} (a_i - a_j)^T (a_i - a_j)\right)}
\end{aligned} \tag{5}$$

For u and v d -vectors, the product of the vector of the form $(u - v)$ with its transpose $(u - v)^T$ is equivalent to the squared Euclidean distance between u and v . Let d_{ij} represent the squared Euclidean distance between a_i and a_j , then the entropy is rewritten as shown in Equation 6, which we observe is congruous with Equation 2 when $d = 4$.

$$H_2(USM) = -\ln \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{(2\pi^{1/2} \sigma)^d} e^{\left(-\frac{1}{4\sigma^2} d_{ij}\right)} \tag{6}$$

4.2 Method 2: Approximate Entropy (ApEn)

In addition to the USM derivation of entropy, we consider two other popular entropy methods for detecting change in complexity of time series which have previously been applied to smart home data. The first of these, Approximate Entropy, is commonly used to compare the complexity of short real-valued sequences. However, it has been shown to perform well for binary symbolic sequences and should in theory be appropriate for alphabets of any size [13].

ApEn requires two parameters to be defined, the embedding dimension m and tolerance r . Given a sequence X of length N and parameters $m = 2$ and $r = 1$, ApEn generates a set of vectors of m values from a sliding window moved over X . For each vector in the set, $x_m(i)$, the maximum element-wise distance (Chebyshev distance) is calculated. Let B_i represent the number of vectors whose Chebyshev distance is $\leq r$. We then calculate the values:

$$C_i^m(r) = B_i / (N - m + 1) \phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{7}$$

Next, the steps are repeated for vectors of size $m + 1$ and A_i , specified as the number of vectors with Chebyshev distance from $x_{m+1}(i) \leq r$. Finally, ApEn(m, r) is defined as $\lim_{N \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]$, which is estimated as:

$$\begin{aligned}
ApEn(m, r, N) &= \phi^m(r) - \phi^{m+1}(r) \\
&\approx (N - m)^{-1} \sum_{i=1}^{N-m} -\ln(A_i/B_i)
\end{aligned} \tag{8}$$

The above statistic is typically approximated by only considering the range $1 \leq i \leq N - m$ for the calculation of A_i and B_i , thus excluding the vector at index $i = N - m + 1$, and allowing the formula to be simplified as shown above.

4.3 Method 3: Sample Entropy (SampEn)

One drawback of ApEn is its bias on super short sequences that causes ApEn to skew to 0, sometimes incorrectly implying sequence regularity [14]. Richman and Moorman [2000] introduced Sample Entropy as a way to reduce this bias. The process of extracting vectors and calculating $A(r)$ and $B(r)$ is performed similarly to

ApEn, except that the match of $x_m(i)$ with itself is not included in the totals, as it is with ApEn and instead of approximating the final value as shown in Equation 8, the value is calculated as the limit ($\lim_{N \rightarrow \infty}$), of the negative logarithm of the ratio between $A^m(r)$ and $B^m(r)$ and the statistic SampEn(m, r, N) as:

$$SampEn(m, r, N) = -\ln [A^m(r)/B^m(r)] \tag{9}$$

5 CONVERGENCE ANALYSIS

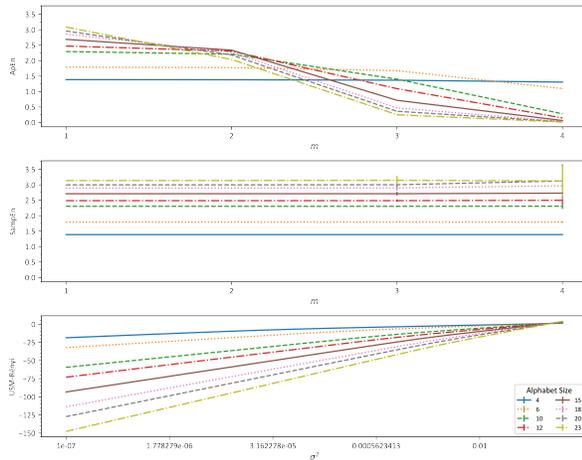
Because we model a person's frailty based on movement entropy, we are interested in determining how well the three proposed methods, USM-Rényi, ApEn, and SampEn, converge and distinguish sequences from distributions of different complexity. We conduct experiments to comparatively evaluate these methods using a Monte Carlo simulation of synthetic symbolic sequence data. All analyses are conducted using Python 3 and source code for USM functions can be found here <https://doi.org/10.5281/zenodo.8180741>. Other functions and data available on request.

We generate 10^3 samples of independent and identically distributed (i.i.d.) uniformly random sequences of length 500 to 10000 for alphabet sizes ranging from 4 to 23. For each sample sequence, we compute ApEn and SampEn for each $m = (1, 2, 3, 4)$ with the tolerance $r = 0$ as r is not needed when dealing with symbolic sequences [13]. We also compute USM coordinates and USM-Rényi entropy with the same kernel variance values, σ^2 , used in [18]. To understand the precision of each entropy measure we estimate the standard error (SE) of its sampling distribution as the standard deviation of the entropy estimates for each sequence length and distribution.

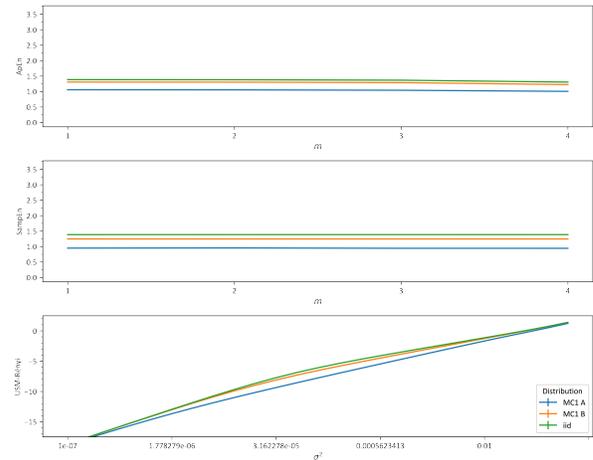
The size of σ^2 in the USM-Rényi estimate roughly corresponds to the length k of k -gram density being estimated [18], therefore there is a congruence between σ^2 and the value of m used in ApEn and SampEn. Pairwise Pearson's correlations between USM-Rényi, ApEn and SampEn for each value of σ^2 and m reveal that while for certain pairs of m and σ^2 there is strong correlation ($|r| \geq 0.9$ for some pairs), the entropy measures are not one to one interchangeable and may each offer unique insight into the complexity of a sequence. However, analysis of the distributions and pairwise Pearson's correlations of the USM-Rényi values reveals for $\sigma^2 < 1 \times 10^{-7}$ $SE \approx 0$ and for $\sigma^2 \geq 0.1$ USM-Rényi estimates are completely correlated with each other ($r = 1$). This implies no new information is gained regarding the complexity of the distribution of USM coordinates from kernel variances beyond these cutoffs and so we limit the remainder of our analyses to σ^2 values in the range $1 \times 10^{-7} \leq \sigma^2 \leq 0.1$.

For all entropy estimates, convergence rates are slower as the alphabet size, d , increases. For $d = 23$ USM-Rényi values for each σ^2 converge for $n \geq 2500$ with small SE relative to the absolute values of the entropy measure. The absolute value of USM-Rényi is consistently greater for larger alphabet size for $1 \times 10^{-7} \leq \sigma^2 \leq 0.01$, though an inflection does occur at $0.0316 < \sigma^2 \leq 0.056$ where the signs of the estimates flip from negative to positive.

One downside to SampEn is it may return an undefined result when there are no $m + 1$ length vector matches. In our simulations, the rate at which SampEn was undefined increased with alphabet size, such that, for $d = 23$, 65% and 19% of SampEn($m=3$) values were undefined for $n = 500$ and $n = 1000$ respectively. And SampEn($m=4$)



(a) Plot of mean entropy measures for i.i.d. sequences length 10000 for each alphabet size. The top two plots show mean ApEn and SampEn against values of m . The bottom plot shows Mean USM-Rényni for each σ^2 . Error bars represent the estimated SE of the sampling distribution. Note that SE are very small for all entropy estimates except SampEn($m=3$) and SampEn($m=4$).



(b) Plot of mean entropy measures for sequences of length 5000 and alphabet 4 for i.i.d. distribution and two first order markov chains of different entropy rates in the limit as $N \rightarrow \infty$. The true entropy rate for MC1 A is 1.059 and MC1 B is 1.308. Error bars not shown as SE are so small they are not visible at this scale.

Figure 3: Ability of entropy estimates to distinguish sequences with different underlying distributions.

was undefined for nearly all 500 and 1000 length sequences and was undefined for 65% and 16% of sequences length 2500 and 5000 respectively. For i.i.d. uniform distributions both ApEn and SampEn should converge to the same value for any m . Our results show this is the case for SampEn but not ApEn. ApEn($m=1$) converges by $n = 2500$ for $d = 23$ but even for sequences length 10000, ApEn convergence is progressively worse for increasing m and d (Figure 3). Despite this slow convergence, for $n \geq 2500$ the spread of ApEn is very narrow with $SE < 0.01$ for all m . SampEn, on the other hand, converges quickly for each value of m even with the smallest sample sizes but the SE for SampEn values is much larger than ApEn to the point of overlapping with other distributions (Figure 3a). For $d = 23$, the SE of SampEn($m=2$) for $n = 2500$ is 0.06, nearly an order of magnitude larger than for ApEn($m=2$) at $n = 2500$ ($SE = 0.008$).

We also use simulated Markov Chains to gauge the ability of the entropy measures to distinguish between sequences with the same alphabet size but different probability distributions. We generate 10^3 sequences of length $n = 5000$ from two first-order Markov Chains with entropy rate, computed following the method in [19], of 1.059 and 1.308 respectively. All three entropy estimates correctly estimate the i.i.d. sequences to be the most complex and MC1 A as the least complex (Figure 3b). While all three methods have narrow enough SEs to distinguish between the distributions, the relative range of ApEn and SampEn is much smaller than USM-Rényni.

6 SMART HOME CASE STUDY

We highlight the use of USM to analyze the complexity of smart home behavior data through a case study. To illustrate the analysis, we use a technique proposed by Almeida and Vinga [3] to generate 2D representations of USMs. In this visualization, each sensor in the home is assigned to one vertex in an equilateral polygon. The

density of points plotted by the USM algorithm creates a fractal-like pattern of sub-polygons (i.e. sub-quadrants in the 4d version shown in Figure 2) representing multi-order transitions between locations. We analyze maps (Figure 4) generated from three days of data from 13 smart home motion detectors placed around the home of an 83-year-old female resident. For comparison, Figure 4a demonstrates how synthetic uniform random data for this smart home configuration generates a map with fully filled nested polygon rings.

The remaining plots in Figure 4 represent three different days of (self-reported) activity: a normal day, a day with housekeeping activities, and a day with a visitor in the home. On the normal day, we observe that the most complete rings occur within a room, with occasional transitions between bedroom and bathroom, living room and sink, and entryway and living room. Overall, transitions between distinct areas of the home are less common.

In contrast, the days with housekeeping and visitor activities generate darker, more complete rings. Figure 4c shows the distribution of motion is fairly uniform for the housekeeping day, while the visitor day highlights more frequent multi-order transitions in and between the living room, kitchen, and dining area. The USM maps align with the reported activities of the participant, visualizing both the amount and diversity of movement within the space, including transitions between regions. This visualization provides a valuable foundation for analyzing movement complexity as an indicator of frailty.

7 DISCUSSION AND CONCLUSION

In this paper, we introduce and compare entropy estimates of smart home-based human behavior as a foundation for analyzing frailty.

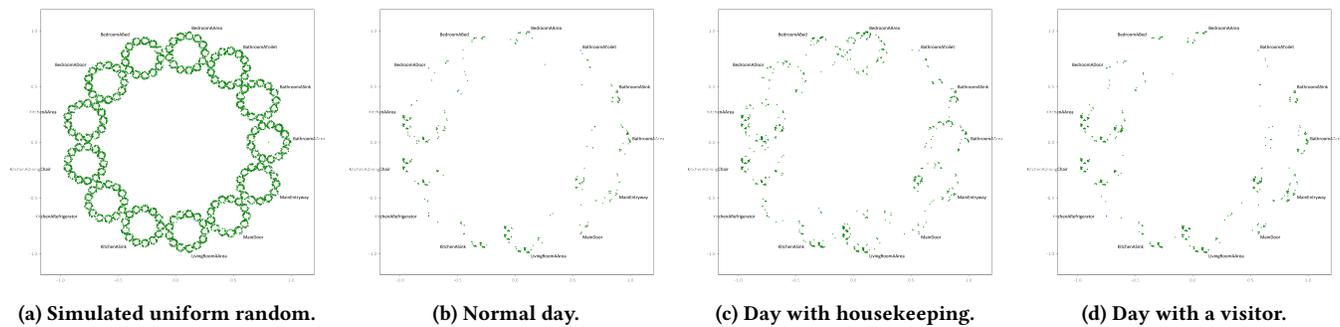


Figure 4: USM plots for 13 motion sensors in a smart home. The completeness of nested sub-polygons indicates the relative frequency of transitions from other locations.

In addition to the popular approximate and sample entropy measures, we also consider universal sequence maps for analyzing such data. We provide a proof of USM based Rényi entropy for any alphabet size. This is a useful advance as USMs provide a simple method to compute the multi-order transition frequencies of smart home behaviors which can be described by an arbitrary number of ambient sensors. USMs also provide a valuable method for visualizing behavior complexity from collected ambient sensor readings.

Our analysis shows that these entropy measures are very dependent on both sequence length and alphabet size. While all methods converge on simulation data, faster convergence also sometimes coexisted with larger error. Researchers should exercise caution when drawing conclusions related to differences in entropy of smart home data if the sequences are very short or of different lengths, or if they come from smart homes with different sensor counts. Of the methods we analyzed, USM-Rényi showed a favorable balance of convergence rate, precision, and relative consistency across different alphabets and distributions compared to ApEn and SampEn.

A limitation of the work is that analysis is based on simulation data. We conjecture that behavior complexity, estimated by entropy of ambient sensor data, is an indicator of frailty. Further work is needed to validate this based on clinical assessment of subject frailty and how it compares with other behavior features such as total within-room and between-room movement. Additionally, our visualizations reflect periods of increased complexity that are due to visitors in the home. Analysis of frailty will need to control for such cases to analyze behavior that is based on one specific person in the home. This work provides a new method to model behavior complexity from smart home sensor data, capturing temporal movement patterns and creating a foundation for passive, continuous in-home health assessment.

ACKNOWLEDGMENTS

The authors would like to thank Aaron Crandall and Brian Thomas for their help with the data collection. This material is based upon work supported by the National Institutes of Nursing Research under Grant No. R01NR016732. This research used resources of the Center for Institutional Research Computing at Washington State University.

REFERENCES

- [1] J S Almeida, J A Carrico, A Maretzek, P A Noble, and M Fletcher. 2001. Analysis of genomic sequences by Chaos Game Representation. *Bioinformatics* 175, 5 (2001), 429–437.
- [2] J S Almeida and S Vinga. 2002. Universal sequence map (USM) of arbitrary discrete sequences. *BMC Bioinformatics* 3 (2002), 6.
- [3] Jonas S Almeida and Susana Vinga. 2009. Biological sequences as pictures: A generic two dimensional solution for iterated maps. *BMC Bioinformatics* 10, 100 (2009), 1–7. <https://doi.org/10.1186/1471-2105-10-100> ISBN: 1471210510.
- [4] I Clay. 2020. The future of digital health. *Digital Biomarkers* 4, 1 (2020), 1–2.
- [5] A A Cohen, L Ferrucci, T Fulop, D Gravel, et al. 2022. A complex systems approach to aging biology. *Nature Aging* 2, 7 (2022), 580–591.
- [6] Diane J Cook, Aaron Crandall, Brian Thomas, and Nayaranan Krishnan. 2013. CASAS: A smart home in a box. *Computer* 46, 7 (2013), 62–69.
- [7] Alfonso Delgado-Bonal and Alexander Marshak. 2019. Approximate Entropy and Sample Entropy: A Comprehensive Tutorial. *Entropy* 21, 6 (May 2019), 541. <https://doi.org/10.3390/e21060541>
- [8] L P Fried, Q L Xue, A R Cappola, L Ferrucci, et al. 2009. Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 64, 10 (2009), 1049–1057.
- [9] Emiel O Hoogendijk, Jonathan Afilalo, Kristine E Ensrud, Paul Kowal, Graziano Onder, and Linda P Fried. 2019. Frailty: implications for clinical practice and public health. *The Lancet* 394, 10206 (Oct. 2019), 1365–1375. [https://doi.org/10.1016/S0140-6736\(19\)31786-6](https://doi.org/10.1016/S0140-6736(19)31786-6)
- [10] Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. 2020. An entropy-based approach for anomaly detection in activities of daily living in the presence of a visitor. *Entropy* 22, 8 (2020), 845. <https://doi.org/10.3390/E22080845>
- [11] H J Jeffrey. 1990. Chaos game representation of gene structure. *Nucleic Acids Research* 18, 8 (1990), 2163–2170.
- [12] L A Lipsitz. 2002. The physiologic basis of functional health and frailty. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 57, 3 (2002), 115–125.
- [13] S Pincus. 2008. Approximate entropy as an irregularity measure for financial data. *Econometric Reviews* 27 (2008), 329–362. Issue 4-6.
- [14] J S Richman and J R Moorman. 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology Heart and Circulatory Physiology* 278 (2000), H2039–H2049.
- [15] K Rockwood and A Mitnitski. 2011. Frailty defined by deficit accumulation and geriatric medicine defined by frailty. *Clinics in Geriatric Medicine* 27, 1 (2011), 17–26.
- [16] N Schütz, S E J Knobel, A Botros, M Single, et al. 2022. A systems approach towards remote health-monitoring in older adults: Introducing a zero-interaction digital exhaust. *NPJ Digital Medicine* 5, 1 (2022), 116.
- [17] S-K Teh, I Rawtaer, and H P Tan. 2022. Predictive accuracy of digital biomarker technologies for detection of mild cognitive impairment and pre-frailty amongst older adults: A systematic review and meta-analysis. *IEEE Journal of Biomedical and Health Informatics* 26, 8 (2022), 1–13.
- [18] S Vinga and J S Almeida. 2004. Rényi continuous entropy of DNA sequences. *Journal of Theoretical Biology* 231, 3 (2004), 377–388.
- [19] Tinghui Wang, Diane J Cook, and Thomas R Fischer. 2023. The indoor predictability of human mobility. *IEEE Transactions on Emerging Topics in Computing* 11, 1 (2023), 182–193.
- [20] Y Wang, A Yalcin, and C VandeWeerd. 2020. An entropy-based approach to the study of human mobility and behavior in private homes. *PLoS ONE* 15, 12 (2020), 1–19.