

# Making the most of quite little: Improving AI training for edge sensor time series

November 25 2022

## Enhancing AI Accuracy for Limited Datasets

Edge artificial intelligence (AI)-based sensors are useful for classification of time series data

But training data for such AI is limited, affecting the classification accuracy of these sensors

### New approach to edge AI training on monitoring cattle behavior

Random sampling of time series snippets expands classifier boundaries

Biased coverage of samples balances out imbalance in original data

Fourier surrogates extend dataset without duplications

**96%** Accuracy enhanced from 90% to 96%

**91%** Grazing classification increased from 45% to 91%

The proposed method can increase edge AI accuracy in behavioral monitoring of farm animals and pave the way to a general methodology for limited datasets

Integrated Data Augmentation for Accelerometer Time Series in Behavior Recognition: Roles of Sampling, Balancing and Fourier Surrogates

Li et al. (2022) | IEEE Sensors Journal | 10.1109/JSEN.2022.3219594



Overview of the proposed data augmentation approach. Credit: Tokyo Institute of Technology

Engineers at the Tokyo Institute of Technology (Tokyo Tech) have demonstrated a simple computational approach for improving the way artificial intelligence classifiers, such as neural networks, can be trained based on limited amounts of sensor data. The emerging applications of the Internet of Things often require edge devices that can reliably classify behaviors and situations based on time series.

However, training data are difficult and expensive to acquire. The proposed approach promises to substantially increase the quality of classifier training, at almost no extra cost.

In recent times, the prospect of having huge numbers of Internet of Things (IoT) sensors quietly and diligently monitoring countless aspects of human, natural, and machine activities has gained ground. As our society becomes more and more hungry for data, scientists, engineers, and strategists increasingly hope that the additional insight which we can derive from this pervasive monitoring will improve the quality and efficiency of many production processes, also resulting in improved sustainability.

The world in which we live is incredibly complex, and this complexity is reflected in a huge multitude of variables that IoT sensors may be designed to monitor. Some are natural, such as the amount of sunlight, moisture, or the movement of an animal, while others are artificial, for example, the number of cars crossing an intersection or the strain applied to a suspended structure like a bridge.

What these variables all have in common is that they evolve over time,

creating what is known as time series, and that meaningful information is expected to be contained in their relentless changes. In many cases, researchers are interested in classifying a set of predetermined conditions or situations based on these temporal changes, as a way of reducing the amount of data and making it easier to understand.

For instance, measuring how frequently a particular condition or situation arises is often taken as the basis for detecting and understanding the origin of malfunctions, pollution increases, and so on.

Some types of sensors measure variables that in themselves change very slowly over time, such as moisture. In such cases, it is possible to transmit each individual reading over a wireless network to a cloud server, where the analysis of large amounts of aggregated data takes place. However, more and more applications require measuring variables that change rather quickly, such as the accelerations tracking the behavior of an animal or the daily activity of a person.

Since many readings per second are often required, it becomes impractical or impossible to transmit the raw data wirelessly, due to limitations of available energy, data charges, and, in remote locations, bandwidth. To circumvent this issue, engineers all over the world have long been looking for clever and efficient ways to pull aspects of data analysis away from the cloud and into the sensor nodes themselves.

This is often called edge [artificial intelligence](#), or edge AI. In general terms, the idea is to send wirelessly not the raw recordings, but the results of a classification algorithm searching for particular conditions or situations of interest, resulting in a much more limited amount of data from each node.

There are, however, many challenges to face. Some are physical and stem from the need to fit a good classifier in what is usually a rather

limited amount of space and weight, and often making it run on a very small amount of power so that long battery life can be achieved.

"Good engineering solutions to these requirements are emerging every day, but the real challenge holding back many real-world solutions is actually another. Classification accuracy is often just not good enough, and society requires reliable answers to start trusting a technology," says Dr. Hiroyuki Ito, head of the Nano Sensing Unit where the study was conducted.

"Many exemplary applications of artificial intelligence such as self-driving cars have shown that how good or poor an artificial classifier is, depends heavily on the quality of the data used to train it. But, more often than not, sensor time series data are really demanding and expensive to acquire in the field. For example, considering cattle behavior monitoring, to acquire it engineers need to spend time at farms, instrumenting individual cows and having experts patiently annotate their behavior based on video footage," adds co-author Dr. Korkut Kaan Tokgoz, formerly part of the same research unit and now with Sabanci University in Turkey.

As a consequence of the fact that training data is so precious, engineers have started looking at new ways of making the most out of even quite a limited amount of data available to train edge AI devices. An important trend in this area is using techniques known as "data augmentation," wherein some manipulations, deemed reasonable based on experience, are applied to the recorded data so as to try and mimic the variability and uncertainty that can be encountered in real applications.

"For example, in our previous work, we simulated the unpredictable rotation of a collar containing an acceleration sensor around the neck of a monitored cow, and found that the additional data generated in this way could really improve the performance in behavior classification,"

explains Ms. Chao Li, doctoral student and lead author of the study.

"However, we also realized that we needed a much more general approach to augmenting sensor time series, one that could in principle be used for any kind of data and not make specific assumptions about the measurement condition. Moreover, in real-world situations, there are actually two issues, related but distinct. The first is that the overall amount of training data is often limited. The second is that some situations or conditions occur much more frequently than others, and this is unavoidable. For example, cows naturally spend much more time resting or ruminating than drinking."

"Yet, accurately measuring the less frequent behaviors is quite essential to properly judge the welfare status of an animal. A cow that does not drink will surely succumb, even though the accuracy of classifying drinking may have low impact on common training approaches due to its rarity. This is called the data imbalance problem," she adds.

The computational research performed by the researchers at Tokyo Tech and initially targeted at improving cattle behavior monitoring offers a possible solution to these problems, by combining two very different and complementary approaches. The first one is known as sampling, and consists of extracting "snippets" of time series corresponding to the conditions to be classified always starting from different and random instants.

How many snippets are extracted is adjusted carefully, ensuring that one always ends up with approximately the same number of snippets across all the behaviors to be classified, regardless of how common or rare they are. This results in a more balanced dataset, which is decidedly preferable as a basis for training any classifier such as a neural network.

Because the procedure is based on selecting subsets of actual data, it is

safe in terms of avoiding the generation of the artifacts which may stem from artificially synthesizing new snippets to make up for the less represented behaviors. The second one is known as surrogate data, and involves a very robust numerical procedure to generate, from any existing time series, any number of new ones that preserve some key features, but are completely uncorrelated.

"This virtuous combination turned out to be very important, because sampling may cause a lot of duplication of the same data, when certain behaviors are too rare compared to others. Surrogate data are never the same and prevent this problem, which can very negatively affect the training process. And a key aspect of this work is that the data augmentation is integrated with the training process, so, different data are always presented to the network throughout its training," explains Mr. Jim Bartels, co-author and doctoral student at the unit.

Surrogate time series are generated by completely scrambling the phases of one or more signals, thus rendering them totally unrecognizable when their changes over time are considered. However, the distribution of values, the autocorrelation, and, if there are multiple signals, the crosscorrelation, are perfectly preserved.

"In another previous work, we found that many empirical operations such as reversing and recombining time series actually helped to improve training. As these operations change the nonlinear content of the data, we later reasoned that the sort of linear features which are retained during surrogate generation are probably key to performance, at least for the application of cow behavior recognition that I focus on," further explains Ms. Chao Li.

"The method of surrogate time series originates from an entirely different field, namely the study of nonlinear dynamics in complex systems like the brain, for which such time series are used to help



distinguish chaotic behavior from noise. By bringing together our different experiences, we quickly realized that they could be helpful for this application, too," adds Dr. Ludovico Minati, second author of the study and also with the Nano Sensing Unit.

"However, considerable caution is needed because no two application scenarios are ever the same, and what holds true for the time series reflecting cow behaviors may not be valid for other sensors monitoring different types of dynamics. In any case, the elegance of the proposed method is that it is quite essential, simple, and generic. Therefore, it will be easy for other researchers to quickly try it out on their specific problems," he adds.

After this interview, the team explained that this type of research will be applied first of all to improving the classification of cattle behaviors, for which it was initially intended and on which the unit is conducting multidisciplinary research in partnership with other universities and companies.

"One of our main goals is to successfully demonstrate high accuracy on a small, inexpensive device that can monitor a cow over its entire lifetime, allowing early detection of disease and therefore really improving not only animal welfare but also the efficiency and sustainability of farming," concludes Dr. Hiroyuki Ito. The methodology and results are reported in a recent article published in the *IEEE Sensors Journal*.

**More information:** Chao Li et al, Integrated Data Augmentation for Accelerometer Time Series in Behavior Recognition: Roles of Sampling, Balancing and Fourier Surrogates, *IEEE Sensors Journal* (2022). [DOI: 10.1109/JSEN.2022.3219594](https://doi.org/10.1109/JSEN.2022.3219594)

Chao LI et al, A Data Augmentation Method for Cow Behavior Estimation Systems Using 3-Axis Acceleration Data and Neural

Network Technology, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* (2021). [DOI: 10.1587/transfun.2021SMP0003](#)

Chao Li et al, Data Augmentation for Inertial Sensor Data in CNNs for Cattle Behavior Classification, *IEEE Sensors Letters* (2021). [DOI: 10.1109/LSENS.2021.3119056](#)

Provided by Tokyo Institute of Technology

Citation: Making the most of quite little: Improving AI training for edge sensor time series (2022, November 25) retrieved 6 May 2024 from <https://techxplore.com/news/2022-11-ai-edge-sensor-series.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.