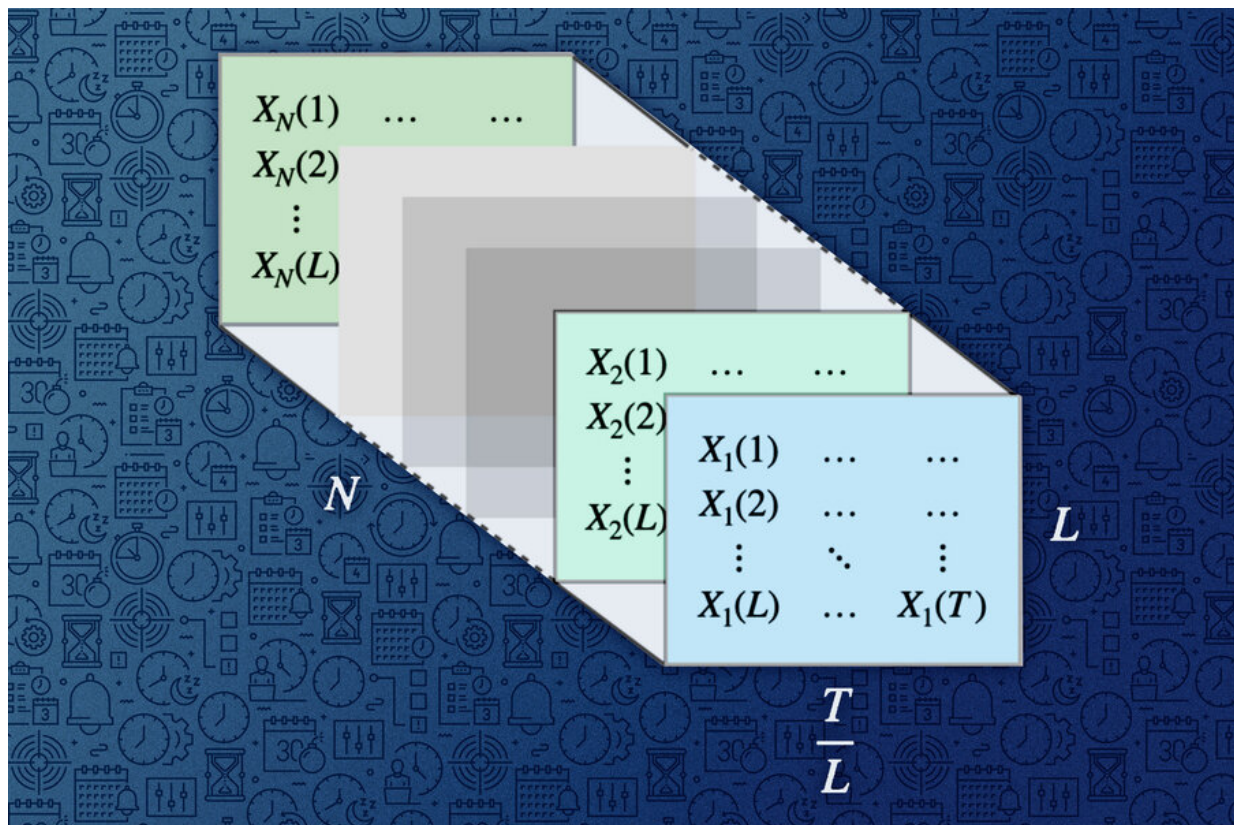# Simplified interface for time-series data predictions

March 28 2022, by Adam Zewe

MIT researchers created a tool that enables people to make highly accurate predictions using multiple time-series data with just a few keystrokes. The powerful algorithm at the heart of their tool can transform multiple time series into a tensor, which is a multi-dimensional array of numbers (pictured). Credit: Figure courtesy of the researchers and edited by MIT News

Whether someone is trying to predict tomorrow's weather, forecast future stock prices, identify missed opportunities for sales in retail, or estimate a patient's risk of developing a disease, they will likely need to interpret time-series data, which are a collection of observations recorded over time.

Making predictions using time-series data typically requires several data-processing steps and the use of complex machine-learning algorithms, which have such a steep learning curve they aren't readily accessible to nonexperts.

To make these powerful tools more user-friendly, MIT researchers developed a system that directly integrates prediction functionality on top of an existing time-series database. Their simplified interface, which they call tspDB (time series predict database), does all the complex modeling behind the scenes so a nonexpert can easily generate a prediction in only a few seconds.

The new system is more accurate and more efficient than state-of-the-art deep learning methods when performing two tasks: predicting future values and filling in missing data points.

One reason tspDB is so successful is that it incorporates a novel time-series-prediction algorithm, explains electrical engineering and computer science (EECS) graduate student Abdullah Alomar, an author of a recent research paper in which he and his co-authors describe the algorithm. This algorithm is especially effective at making predictions on multivariate time-series data, which are data that have more than one time-dependent variable. In a weather database, for instance, temperature, dew point, and cloud cover each depend on their past values.

The algorithm also estimates the volatility of a multivariate time series to

provide the user with a confidence level for its predictions.

"Even as the time-series data becomes more and more complex, this algorithm can effectively capture any time-series structure out there. It feels like we have found the right lens to look at the model complexity of time-series data," says senior author Devavrat Shah, the Andrew and Erna Viterbi Professor in EECS and a member of the Institute for Data, Systems, and Society and of the Laboratory for Information and Decision Systems.

Joining Alomar and Shah on the paper is lead author Anish Agrawal, a former EECS graduate student who is currently a postdoc at the Simons Institute at the University of California at Berkeley. The research will be presented at the ACM SIGMETRICS conference.

Adapting a new algorithm

Shah and his collaborators have been working on the problem of interpreting time-series data for years, adapting different algorithms and integrating them into tspDB as they built the interface.

About four years ago, they learned about a particularly powerful classical algorithm, called singular spectrum analysis (SSA), that imputes and forecasts single time series. Imputation is the process of replacing missing values or correcting past values. While this algorithm required manual parameter selection, the researchers suspected it could enable their interface to make effective predictions using time series data. In earlier work, they removed this need to manually intervene for algorithmic implementation.

The algorithm for single time series transformed it into a matrix and utilized matrix estimation procedures. The key intellectual challenge was how to adapt it to utilize multiple time series. After a few years of

struggle, they realized the answer was something very simple: "Stack" the matrices for each individual time series, treat it as a one big matrix, and then apply the single time-series algorithm on it.

This utilizes information across multiple time series naturally—both across the time series and across time, which they describe in their new paper.

This recent publication also discusses interesting alternatives, where instead of transforming the multivariate time series into a big matrix, it is viewed as a three-dimensional tensor. A tensor is a multi-dimensional array, or grid, of numbers. This established a promising connection between the classical field of time series analysis and the growing field of tensor estimation, Alomar says.

"The variant of mSSA that we introduced actually captures all of that beautifully. So, not only does it provide the most likely estimation, but a time-varying confidence interval, as well," Shah says.

The simpler, the better

They tested the adapted mSSA against other state-of-the-art algorithms, including deep-learning methods, on real-world time-series datasets with inputs drawn from the electricity grid, traffic patterns, and financial markets.

Their algorithm outperformed all the others on imputation and it outperformed all but one of the other algorithms when it came to forecasting future values. The researchers also demonstrated that their tweaked version of mSSA can be applied to any kind of time-series data.

"One reason I think this works so well is that the model captures a lot of time series dynamics, but at the end of the day, it is still a simple model.

When you are working with something simple like this, instead of a neural network that can easily overfit the data, you can actually perform better," Alomar says.

The impressive performance of mSSA is what makes tspDB so effective, Shah explains. Now, their goal is to make this algorithm accessible to everyone.

Once a user installs tspDB on top of an existing database, they can run a prediction query with just a few keystrokes in about 0.9 milliseconds, as compared to 0.5 milliseconds for a standard search query. The confidence intervals are also designed to help nonexperts to make a more informed decision by incorporating the degree of uncertainty of the predictions into their decision making.

For instance, the system could enable a nonexpert to predict future stock prices with high accuracy in just a few minutes, even if the time-series dataset contains missing values.

Now that the researchers have shown why mSSA works so well, they are targeting new algorithms that can be incorporated into tspDB. One of these algorithms utilizes the same model to automatically enable change point detection, so if the user believes their time series will change its behavior at some point, the system will automatically detect that change and incorporate that into its predictions.

They also want to continue gathering feedback from current tspDB users to see how they can improve the system's functionality and user-friendliness, Shah says.

"Our interest at the highest level is to make tspDB a success in the form of a broadly utilizable, open-source system. Time-series data are very important, and this is a beautiful concept of actually building prediction

functionalities directly into the database. It has never been done before, and so we want to make sure the world uses it," he says.

"This work is very interesting for a number of reasons. It provides a practical variant of mSSA which requires no hand tuning, they provide the first known analysis of mSSA, and the authors demonstrate the real-world value of their [algorithm](#) by being competitive with or out-performing several known algorithms for imputations and predictions in (multivariate) time series for several real-world data sets," says Vishal Misra, a professor of computer science at Columbia University who was not involved with this research. "At the heart of it all is the beautiful modeling work where they cleverly exploit correlations across time (within a time series) and space (across time series) to create a low-rank spatiotemporal factor representation of a multivariate time series. Importantly this model connects the field of time series analysis to that of the rapidly evolving topic of tensor completion, and I expect a lot of follow-on research spurred by this paper."

More information: Anish Agarwal, Abdullah Alomar, Devavrat Shah, On Multivariate Singular Spectrum Analysis and its Variants. arXiv:2006.13448v3 [cs.LG], arxiv.org/abs/2006.13448

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology