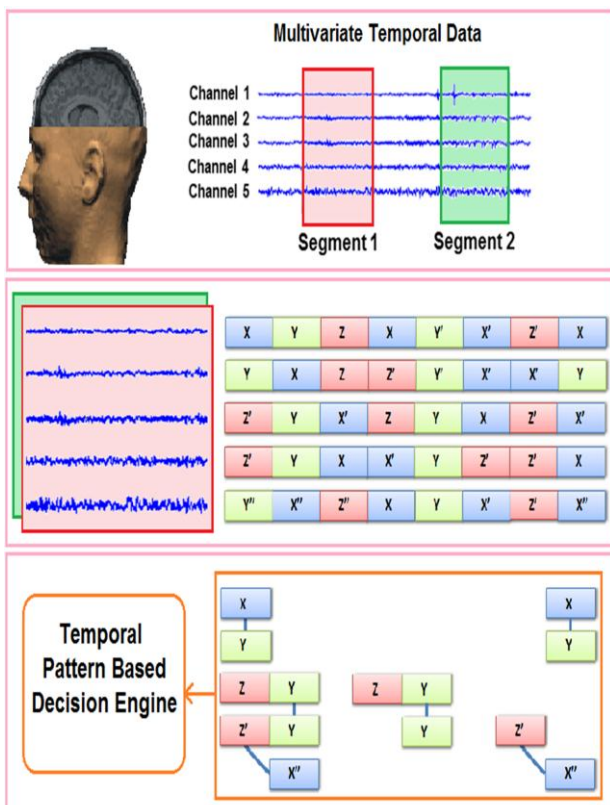


NSF Highlights

Temporal Pattern Discovery in High-throughput Data Domain



A temporal pattern based classification system for multivariate time series data in high-throughput data disciplines: Multivariate temporal data is extracted from the human brain, and then divided into small segments, Segments 1 and 2 in the image. These segments are converted into temporal events, shown as x, y, and z. Electrodes are represented by Channels 1-5.

Credit: Sumeet Dua, Sheetal Saini, LaTech
Image provided by sdua@coes.latech.edu

Multivariate temporal data are collections of contiguous data values that reflect complex temporal changes over a given duration. Technological advances have resulted in huge amounts of such data in high-throughput disciplines, including data from MD simulations of biology, physics, and chemistry, EEG and iEEG data in health-informatics, and stock market data in finance. Traditional data analysis and data-mining techniques are effective in capturing global trends, but fail to capture localized behavioral changes in large temporal data sets. Therefore, specialized temporal data-mining techniques are needed to extract localized temporal patterns. This methodology will aid researchers in overcoming the limitations of conventional techniques by providing previously unknown, nontrivial, and potentially useful temporal patterns as easy to interpret temporal relations that are otherwise unavailable through conventional data analysis techniques. Thus, a multivariate temporal pattern decision engine has been created for this initiative.

The project, spearheaded by Dr. Sumeet Dua, Endowed Associate Professor of Computer Science and his research team at LA Tech, has resulted in a unique algorithm that efficiently and accurately mines high-throughput time series data to extract temporal patterns of interest. These temporal patterns are further utilized to classify multivariate temporal data. For this task, a novel data-mining algorithm has been developed to identify temporal patterns of interest in multivariate data.¹ These temporal patterns are then used to design an automated system to classify multivariate time series data. The algorithm is based on the hypothesis that unique temporal patterns of interest can be extracted from data and that these patterns have significant discriminatory power. This hypothesis has been validated using iEEG data for epileptic seizure detection and can be extended to other disciplines.

Multivariate Temporal Pattern Decision Engine: The developed classification system is a novel, data-mining-based, automated system that classifies multivariate time series data. The proposed system is intended as a generic classification framework. Although, multivariate iEEG time series data is used for epileptic seizure detection to validate the algorithm, it can be extended to analyze other time series data that are generated in other disciplines.

Primary Strategic Outcome Goal/Discovery

Experimental Program to Stimulate Competitive Research (EPSCoR)

Computer & Information Science and Engineering

Cyberinfrastructure

¹S. Dua, S. Saini and H. Singh, "Temporal Pattern Mining for Multivariate Time Series Classification," *Journal of Medical Imaging and Health Informatics* (in press), 2010.

NSF Highlights

Temporal Pattern Discovery in High-throughput Data Domain

Computational Biology

Secondary Strategic Outcome Goals:

Does this highlight represent potentially transformative research? so please explain why. For more information, see [Report to Congress: Transformative Research at the National Science Foundation, April, 2008](#) and [Important Notice 130: Transformative Research](#).

Yes

A novel temporal data mining based automated system is developed for epileptic seizure detection

What is the intellectual merit of this activity?

The developed automated system provides a novel and versatile classification system that can be applied to classify high-throughput time series data irrespective of the discipline in which the data is generated. This study further reinforces the hypothesis that temporal patterns are indeed identifiable from the time series data.

What are the broader impacts of this activity?

[Merit Review Broader Impacts Criterion: Representative Activities, July 2007](#)

- Integrate research with education activities in order to communicate in a broader context

The project has led to a renewed relationship with researchers from the medical and clinical areas fuelling further interdisciplinary and inter-institutional opportunities for research, development and education.