An Introduction to Similarity Analysis Using SAS®

Michael Leonard, SAS Institute Inc.
Jennifer Lee, SAS Institute Inc.
Taiyeong Lee, SAS Institute Inc.
Bruce Elsheimer, SAS Institute Inc.

ABSTRACT

Web sites and transactional databases collect large amounts of time-stamped data related to an organization's suppliers and/or customers over time. Mining these time-stamped data can help business leaders make better decisions by enabling them to listen to their suppliers or customers via their transactions collected over time. A business can have many suppliers and/or customers and might have a set of transactions associated with each one. However, each set of transactions might be quite large, making it difficult to perform many traditional data mining tasks. This paper proposes techniques for large-scale similarity analysis using similarity measures combined with automatic time series analysis and decomposition. After similarity analysis, traditional data mining techniques can then be applied to the similarity analysis results along with other profile data. This paper demonstrates these techniques using SAS/ETS®, SAS/STAT®, SAS® Enterprise Miner™, and SAS® High-Performance Forecasting software.

No paper was submitted for publication.

CONTACT INFORMATION

Michael Leonard
SAS Institute Inc.
Michael.Leonard@sas.com