Decision Support Systems 113 (2018) 99–107

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Visual interactive support for selecting scenarios from time-series ensembles

Guilherme G. Schardong*, Ariane M.B. Rodrigues, Simone D.J. Barbosa, Hélio Lopes

Departamento de Informática, PUC-Rio, Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro, RJ 22451-900, Brazil

ARTICLE INFO

Keywords:
Scenario reduction
User interaction
Time series
Multidimensional projection
Ensemble data

ABSTRACT

Stochastic programming approaches to solve the scenario reduction problem have become invaluable in the analysis and behavior prediction of dynamic systems. However, such techniques often fail to take advantage of the user's own expertise about the problem domain. This work provides visual interactive support to assist users in solving the scenario reduction problem with time-series data. We employ a series of time-based visualization techniques linked together to perform the task. By adapting a multidimensional projection algorithm to handle temporal data, we can graphically present the evolution of the ensemble. We also propose to use cumulative bump charts to visually compare the ranks of distances between the ensemble time series and a baseline series. To evaluate our approach, we developed a prototype application and conducted observation studies with volunteer users of varying backgrounds and levels of expertise. Our results indicate that a graphical approach to scenario reduction may result in a good subset of scenarios and provides a valuable tool for data exploration in this context. The users liked the interaction mechanisms provided and judged the task to be easy to perform with the tools provided.

1. Introduction

Recent developments in simulation techniques have helped researchers to better understand and predict several naturally occurring phenomena, ranging from weather forecast [1] to circuit calibration [2] and fluid dynamics [3]. These simulations produce a huge amount of data, due to the availability of computing power and simulation model refinements. To extract meaningful information from all those data, researchers have been developing an array of approaches in diverse areas: data mining [4], machine learning [5], visualization [6], and optimization [7]. Such approaches typically use statistical measures to summarize the results or to reduce the dimensionality of data and select the most probable outcomes of the simulation.

In simulation analysis, scenario reduction is particularly useful, since its goal is to reduce the number of simulation outcomes (i.e., *scenarios*) to a more manageable size, with minimal loss of variability. Existing approaches are usually modeled as stochastic programming problems [8,9], in which a probability is associated to each possible scenario, and the goal is to select a subset of scenarios whose probability is closest to that of the original set (henceforth called *ensemble*). However, none of those approaches actively engages the users and their knowledge about the problem domain. Visualization-based approaches, conversely, allow for interactive exploration of the data through visual tools and interaction mechanisms. Visual analytics supports decision making by integrating the best of computational processing power and human cognitive prowess [10-13]. For time-based ensembles, Cheng et al. [14] provide a comprehensive survey on time-series and time-based visualization techniques and interaction mechanisms, from which we draw in our proposal.

The main goal of our work is to provide visual interactive support for solving the scenario reduction problem with time-series data. We employ a series of time-based visualization techniques linked together, allowing the user to draw from the strengths of each technique. To the best of our knowledge, no one has proposed a similar way of approaching this problem.

We also propose adaptations to two known visualization algorithms: (i) the Local Affine Multidimensional Projection (LAMP) algorithm [15], in order to produce a time-based representation of the data; and (ii) Bump charts, also known as Slope graphs [16], in order to view a transformed version of our time-series ensemble. Multidimensional projections are evolving, especially in the works of Alencar et al. [17] and Wong et al. [18]. By using a different base technique, with a strong mathematical foundation [15], we aim to provide a more robust representation of the similarity between the series over time. We have also made some adaptations to Bump charts [16]. As our data is not

* Corresponding author.
E-mail addresses: gschardong@inf.puc-rio.br (G.G. Schardong), arodrigues@inf.puc-rio.br (A.M.B. Rodrigues), simone@inf.puc-rio.br (S.D.J. Barbosa), lopes@inf.puc-rio.br (H. Lopes).

https://doi.org/10.1016/j.dss.2018.08.001
Received 9 February 2018; Received in revised form 21 June 2018; Accepted 2 August 2018
Available online 03 August 2018

ordinal, we transform them by ranking the distance between each series and a baseline. Moreover, we do not treat each time step as isolated from the others, but as an accumulated rank from all previous time steps.

As proof of concept, we built a prototype software using the brushing and linking framework, proposed by Becker et al. [19], Buja et al. [20], as basis for the user interaction with the different visualizations. We chose four visualization mechanisms: (i) a Fanchart, proposed by Britton et al. [21]; (ii) the Distance scatterplot; (iii) a cumulative Bump chart; and (iv) a scatterplot with the results of our proposed multidimensional projection (MP) approach. To evaluate our approach, we conducted an empirical study involving experts and non-experts in the scenario reduction area.

In summary, our contributions are:

- A visual interactive approach to assist the user in selecting a subset of meaningful scenarios from a time-series ensemble dataset, thus solving an instance of the scenario reduction problem;
- An adaption of a multidimensional projection algorithm to generate a visual representation of time-varying data, taking into account the time component of the data;
- A transformation of a time-series ensemble dataset into a cumulative, ranked version, in order to support a visual assessment of its evolution.

The remainder of this paper is organized as follows: Section 2 presents related works on Scenario Reduction and Multidimensional Projection. In Section 3 we describe our approach and explain the visualization techniques employed. Section 4 presents the user study plan. Section 5 presents the results of our study and discusses its implications. Finally, Section 6 presents some concluding remarks and directions for future work.

2. Related work

This section describes two groups of related works, on: (i) scenario reduction (i.e., the problem we address); and (ii) multidimensional projections, which are the type of algorithm we have adapted to help us with scenario reduction. Table 1 presents some of the main works and their contributions compared to ours. Details about these works are presented below.

2.1. Scenario reduction

As the number of objects in an ensemble grows, it becomes increasingly difficult to analyze or visualize it adequately, even when using dimensionality reduction techniques. In many cases, it becomes necessary to select a representative subset of the ensemble for further processing, in a process known as *scenario reduction*, which has increasingly attracted researchers' interest, especially in areas such as power production [9,23] and geostatistics [2,8,22,30,33-35]. A number of researchers have proposed to use stochastic programming as an approach to tackle this problem. Dupačová et al. [9] stated that the Fortet-Mourier family of probability metrics may be used as canonical metrics to find a subset of scenarios with probability distributions closest to the original set. They reduced the number of possible scenarios by 50% while keeping 90% relative accuracy in the remaining scenarios. More recently, Armstrong et al. [8] proposed a metric for the distance between conditional simulated realizations of ore deposits, along with a random search procedure to find an approximation of the ideal subset of scenarios. They followed the approach proposed by Heitsch and Römisch [34] to calculate the distance between a subset of scenarios and the full ensemble. In their experiments, the best subset found was 1% off the expected value for their objective function, which indicates that the number of possible scenarios can be strongly reduced without significant loss in variability.

In the petroleum field, Scheidt and Caers [30] have used dimensionality reduction and kernel methods to quantify the uncertainty in an ensemble of geological facies realizations. Their approach involved mapping the realizations onto a lower dimensional space using a multidimensional scaling (MDS) algorithm [36] and flow-related distance metrics, such as the Hausdorff distance [37] or time-of-flight-based metrics [38]. They have also used kernel methods to transform the projected points from a non-linear space onto a linear one, thus facilitating the application of grouping approaches, such as clustering algorithms and Principal Component Analysis. After defining the realization groups, a few elements of each group are chosen for the actual flow simulation. The flow simulation statistics they obtained with a reduced number of realizations were very similar to those with the full ensemble.

Different from most of the scenario reduction approaches presented so far, our main goal is to allow users to input their own knowledge of the problem domain into the process through graphical tools, therefore leading to a more flexible process overall. To the best of our knowledge, there is little work on visual analytics and scenario reduction. Sahaf et al. [31] proposed a scenario reduction approach based on randomly sampling scenarios after clustering them using a mutual information similarity metric. They implemented this approach in a visual analytics framework, where it is possible to visualize the spatial contribution of each model to the similarity of scenarios and run a clustering algorithm in an area specified by the user. Kawas et al. [27] proposed an uncertainty-aware framework for decision optimization, in which they employed classic stochastic programming to perform the scenario reduction and used visual analytics only to evaluate the resulting models.

On a decision support systems context we found no works that involve both scenario reduction and visual analytics. Waser et al. [32] come close, by proposing a scenario generation and interactive visualization approach applied to flooding management. Their approach can simulate flooding scenarios in real time, but their visualization tool is not scalable and the plans generated are suboptimal. Domenica et al. [26] incorporate stochastic programming and scenario generation techniques into established decision support and information systems. They successfully argue that decision and simulation models can be combined in order to create business analytics, therefore creating uncertainty-aware decision and information systems. Park et al. [28] proposed a visual analytics approach for managing supply chain networks. They modeled these networks as directed graphs and implemented a series of interactive visualizations for them, including: force-directed layout, treemap layout, substrate-based visualization, chord diagrams and matrix layout. However, their views are not connected to each other, therefore lacking an important pattern-discovery mechanism.

Table 1
Overview of the literature in scenario reduction and multidimensional projections.

Work	Scenario reduction	Multidimensional projections	Brushing & linking
Armstrong et al. [22]	X		
Gröwe-Kuska et al. [23]	X		
Meira et al. [24]	X		
Lee et al. [2]	X		
Heitsch and Römisch [25]	X		
Domenica et al. [26]	X		
Kawas et al. [27]	X		
Park et al. [28]			X
Demir et al. [29]			X
Scheidt and Caers [30]	X	X	
Sahaf et al. [31]	X		X
Waser et al. [32]	X		X
Our approach	X	X	X