Pattern Discovery: A Progressive Visual Analytic System Design to Support Categorical Data Analysis

Hanqing Zhao , Huijun Zhang , Yan Liu , Yongzhen Zhang , Xiaolong (Luke) Zhang

PII: S1045-926X(16)30173-2
DOI: 10.1016/j.jvlc.2017.05.004
Reference: YJVLC 790

To appear in: *Journal of Visual Languages and Computing*

Received date: 19 September 2016
Revised date: 18 April 2017
Accepted date: 20 May 2017

Please cite this article as: Hanqing Zhao , Huijun Zhang , Yan Liu , Yongzhen Zhang , Xiaolong (Luke) Zhang , Pattern Discovery: A Progressive Visual Analytic System Design to Support Categorical Data Analysis, *Journal of Visual Languages and Computing* (2017), doi: 10.1016/j.jvlc.2017.05.004

Pattern Discovery: A Progressive Visual Analytic System Design to Support Categorical Data Analysis

Hanqing Zhao is with the Pennsylvania State University. E-mail: hanqingzhao@psu.edu.

Huijun Zhang is with Taiyuan University of Technology. E-mail: zhanghuijun0038@link.tyut.edu.cn Yan Liu is with the Pennsylvania State University. E-mail: yzl166@psu.edu.

Yongzhen Zhang is with Shanxi Tumor Hospital, China. E-mail:zyzzhang2003@163.com.

Xiaolong (Luke) Zhang is with the Pennsylvania State University and Taiyuan University of Technology. E-mail: xiaolong.zhang@gmail.com.

Abstract—When using data-mining tools to analyze big data, users often need tools to support the understanding of individual data attributes and control the analysis progress. This requires the integration of data-mining algorithms with interactive tools to manipulate data and analytical process. This is where visual analytics can help. More than simple visualization of a dataset or some computation results, visual analytics provides users an environment to iteratively explore different inputs or parameters and see the corresponding results. In this research, we explore a design of progressive visual analytics to support the analysis of categorical data with a data-mining algorithm, Apriori. Our study focuses on executing data mining techniques step-by-step and showing intermediate result at every stage to facilitate sense-making. Our design, called Pattern Discovery Tool, targets for a medical dataset. Starting with visualization of data properties and immediate feedback of users' inputs or adjustments, Pattern Discovery Tool could help users detect interesting patterns and factors effectively and efficiently. Afterward, further analyses such as statistical methods could be conducted to test those possible theories.

Index Terms—Progressive, visual analytics, categorical data analysis

1 Introduction

Facing the challenge of 'Big Data', analysts often use analytic tools to support the understanding of analysis process, including individual data attributes and control of analysis steps. The integration of data-mining algorithms with interactive tools is required to manipulate data and control analysis process. Visual analytics techniques can help to achieve this goal. In traditional visual analytic systems, users typically give some inputs and wait for the processing of a whole dataset until the updated result comes out. However, this approach is inadequate when complex data mining techniques are involved in visual analytics [1, 2]. First, large data volumes would result in long response time and slow down user's exploration of datasets. Second, traditional analytics workflows treat computation processes as a black box. With today's complex data mining techniques, users become more isolated from the computation process, which would result in low situation awareness and hinder users from making efficient and effective decisions.

Different approaches have been proposed to solve the problems above. Liu et al. [3] provided a thorough survey of developments in high-dimensional data visualization over the past decade. Following the principle of facilitating exploration, progressive visualization was