Visual Analytics of Relations of Multi-Attributes in Big Infrastructure Data

Jianlong Zhou, Zelin Li, Zongjian Zhang, Bin Liang, and Fang Chen
Data61, CSIRO
13 Garden Street, Eveleigh
NSW 2015, Australia
Email: FirstName.LastName@data61.csiro.au

***Abstract*—This paper presents information visualization methods for revealing relations of multi-attributes in big infrastructure data. The interactive parallel coordinates, sunburst visualization and combinational visualization approaches are used to represent different relations to get insights from the big infrastructure data. The water pipe failure data is used as a case study to show the effectiveness of proposed visual analytics approaches.**

***Index Terms*—Visual analytics, relations, multi-attributes.**

I. Introduction

With the rapid increase of data from various fields such as biology, finance, medicine, and society, users are looking to integrate their "Big Data" and advanced analytics into business operations in order to become more analytics-driven in their decision making. Besides the development of Machine Learning (ML) models to get insights from big data, visual analytics has been increasingly becoming a critical tool to let users interactively analyze and get understandable data with the use of human cognitive and perceptual principles [1].

For the most real-world big data, one of the important information that helps users understand data and make further decisions is the relations between different attributes in the data. Using the big infrastructure data, such as water pipe data, water supply networks constitute one of the most crucial and valuable urban assets. The combination of growing populations and aging pipe networks requires water utilities to develop advanced risk management strategies in order to maintain their distribution systems in a financially viable way [2]. Especially for critical water mains (generally >300 mm in diameter), defined based on the network location (for example, a single trunk line connecting distribution areas or under a major road) or size which infers impact potential, failures of them typically bring severe consequences due to service interruptions and negative economic and social impacts, such as flooding and traffic disruption [2]. The financial and social costs of reactive repairs in such scenarios amount to more than one billion dollars annually in Australia alone.

If high-risk pipes can be identified before a failure occurs, it is likely that repairs can be completed with minimal service interruption, water loss and negative reputational and community impacts. Identification of an accurate predictor measure that indicates imminent failure will allow utility companies to take actions to mitigate the failure for a lower cost than repairing a full-scale failure. This will contribute to extending the service life of pipes that are still in good condition and allow to run the mains to an acceptable defined risk limit [2]. Various ML models have been developed to predict water pipe failures based on historical water pipe failure records [2]. However, besides the failure predictions, domain experts are also interested in understanding how much different pipe attributes (e.g. the pipe age and pipe material) contribute to pipe failures as well as relations among different attributes in order to make informative decisions.

978-1-5090-5272-1/16$31.00 © 2016 Crown

This paper proposes different information visualization methods to reveal relations of multiple attributes in infrastructure data. Such visual representations benefit both domain experts and data scientists for better understanding contributions of different attributes to target values (e.g. pipe failures). In summary, the contributions of the paper include: 1) interactive visual analytics of relations between multiple attributes and target variable with the use of parallel coordinates based approach, 2) visualization of relations between different attributes with the use of sunburst visualizations to understand the overall pipe attributes, and 3) the combinational color-encoded visualization of relations among multiple attributes to understand exact distributions of various attributes in the data.

II. Visual Analytics of Relations

A. Information Available for Problems

The management authorities of water pipes collect various data on water pipes, which include geographical information of pipes, failure history, and attributes of pipes. Specifically, these data mainly include:

- Geographical and environmental factors: location of pipes, soil types, weather conditions and transportation conditions around the pipe area;
- Failure history: failure type, failure date, failure times;
- Pipe attributes: laid year, size (diameter), length, materials (typically include cement mortar lined cast iron, ductile iron or steel, asbestos cement, or plastic), coating types;
- Internal pressure.

This paper focuses on the visualization of pipe attributes of laid year, diameter size, material, failure year and predicted failure probability by machine learning models.

B. Relations of Multi-Attributes

Relation Revealing with Parallel Coordinates One of classical approaches to visualize multi-attribute data is parallel coordinates [3]. It can provide an overview of data trend. In this paper, we propose a visualization approach for presenting multi-attributes by color-encoded data points in parallel coordinates. In this visualization, each vertical axis represents one data attribute of predicted failures (target variable). Each pipe is represented with a polyline connecting attribute values on each vertical axis. All pipes belonging to same/similar values of a specific attribute are encoded with colors. Such color encoding approach makes pipes belonging to a common or similar data range of an attribute easily differentiable, which provides an overview of data trend of pipes. Such benefits of providing an overview of pipes and their associated attribute details can improve the information browsing efficiency. Fig. 1 shows an example of visualization of relations of multi-attributes of pipes and failure probability with the parallel coordinates. Fig. 2 shows relations of a specific range of an attribute with other attributes. It shows that for the pipes laid in the year around 1900, the material of all pipes is the CICL. Their diameter size ranges from 100mm to 900mm. They failed spreading from 2001 to 2012 and their predicted failure probability ranges from 0.04 to 0.20.

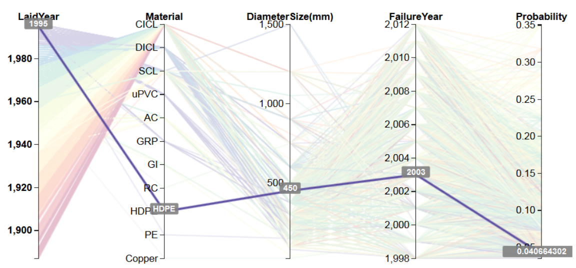

Fig. 1. Visualization of relations of multi-attributes and target variable with the parallel coordinates.

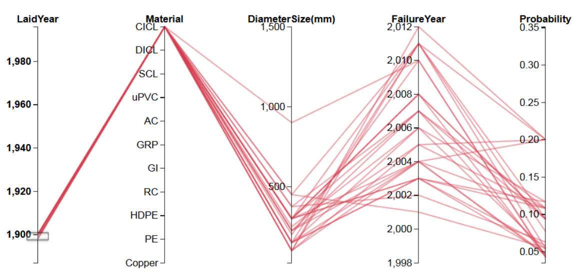

Fig. 2. Relations of a specific range of laid year of pipes with other attributes.

Relation Revealing with Sunburst Visualization Fig. 3 shows the visualization of relations between different attributes with the use of sunburst visualizations. Such visualization helps users understand the relations of overall pipe attributes in the data set.

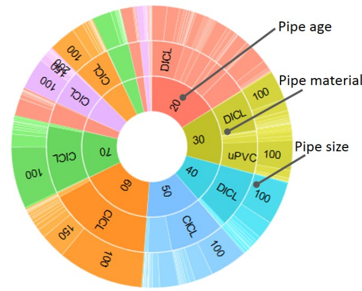

Fig. 3. Visualization of overall pipe attribute distributions.

Combinational Visualization of Relations of Multi-Attributes Fig. 4 uses combinational visualization of linked table, pie chart and bar chart to reveal relations among multiple attributes. It allows users to interactively inspect multi-attribute relations with more details.

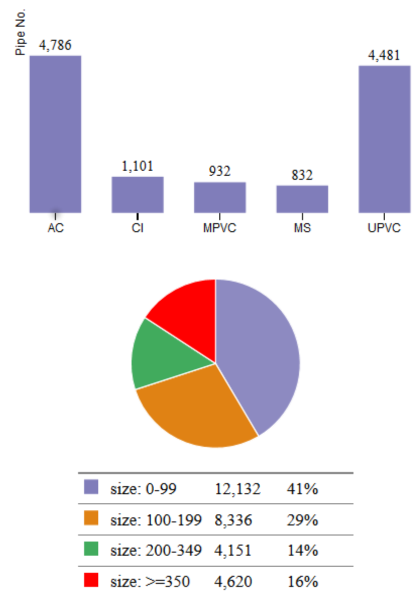

Fig. 4. The combinational visualization of relations among multi-attributes.

III. Conclusions and Future Work

This paper proposed different visual analytics approaches to reveal relations of multi-attributes in big infrastructure data. We used water pipe failure data as a case study to show the effectiveness of proposed approaches. Our future work will focus on the visual analytics of contributions of different attributes to prediction results from machine learning models.

References

[1] J. Zhang, E. Yanli, J. Ma, Y. Zhao, B. Xu, L. Sun, J. Chen, and X. Yuan, "Visual analysis of public utility service problems in a metropolis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1843–1852, 2014.

[2] Z. Li, B. Zhang, Y. Wang, F. Chen, R. Taib, V. Whiffin, and Y. Wang, "Water pipe condition assessment: A hierarchical beta process approach for sparse incident data," *Machine Learning*, vol. 95, no. 1, pp. 11–26, 2014.

[3] J. Zhou, C. Xiao, Z. Wang, and M. Takatsuka, "A concept of volume rendering guided search process to analyze medical data set," *Computerized Medical Imaging and Graphics*, vol. 32, no. 2, pp. 140–149, 2008.