Journal of Visual Languages and Computing 47 (2018) 19–30

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

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

A Multi-comparable visual analytic approach for complex hierarchical data

Chen Yi[a,*], Dong Yu[a], Sun Yuehong[a], Liang Jie[b]

[a] Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China

[b] Faculty of Engineering and IT, School of Software, University of Technology Sydney, New South Wales, Australia

ARTICLE INFO

Keywords:
Visual analysis
Hierarchy comparison
Multi-dimensional data
Evaluation metric
MRL standard

ABSTRACT

Maximum residue limit (MRL) standard which specifies the highest level of every pesticide residue in different agricultural products plays a critical role in food safety. However, such standards which related to the characteristics of pesticides and the classification of agricultural products which organized into a hierarchical structure are complex and vary widely across different regions or countries. So it is a big challenge to compare multi-regional MRL standard data comprehensively. In this paper, we present a multi-comparable visual analytic approach for complex hierarchical data and a visual analytics system (McVA) to support multiple comparison and evaluation of MRL standard. With a cooperative multi-view visual design, our proposed approach links the hierarchies of MRL datasets and provides the capacity for comparison at different levels and dimensions. We also introduce a metric model for evaluating the completeness and strictness of MRL standards quantitatively. The case study of real problems and the positive feedback from domain experts demonstrate the effectiveness of this approach.

1. Introduction

Comparison of two complex hierarchical datasets is very important in many fields. For example, Maximum Residue Limit (MRL) standard which specifies the highest level of every pesticide residue in different agricultural products plays a critical role in food safety. It is necessary to compare one country's MRL standard with other's in order to revise and supplement the standard. However, such standards which related to the characteristics of pesticides and the classification of agricultural products which organized into a hierarchical structure are complex and vary widely across different countries or regions. So, it is a big challenge and a tough work to compare two MRL standards comprehensively.

In this paper, we present a multi-comparable visual analytic approach for comparison and evaluation of complex hierarchical data, and further implement an interactive visual analytics system with a cooperative multi-view design named McVA shown in Fig. 1. McVA links multi-regional MRL datasets and provides the capacity for comparison at different levels and dimensions. The main contributions of this work are as follows:

(1) A multi-comparable visual analytic approach is presented to compare complex hierarchical datasets by a suite of data processing and visualization methods. After data processing, multiple hierarchical structures are abstracted from the complex dataset. Nested circles with lasso tools were used to compare hierarchical structure interactively, parallel coordinates encoded by different classifications and colors were used to analysis multi-dimensional attributes interactively.

(2) A new evaluation model for comparison of complex hierarchical datasets is introduced. It includes three metrics: (a) Dissimilarity degree of nodes' attribute value in homogenous part, (b) Diversity of structures in heterogeneous parts, and (c) The comprehensive dissimilarity degree of two complex data sets. All these three metrics can be used to compare two hierarchical data sets quantitatively in both selected and total structure.

(3) A visual analytic system (McVA) is implemented for comparative analysis of multidimensional and hierarchical data. McVA provides a suite of visualization designs and selecting, filtering and interaction tools to help user for comparison of MRL standards from different countries or regions. It also can help users to explore the similarity and difference of homogenous part, heterogonous part and the whole dataset quantitatively.

* Corresponding author.
E-mail address: chenyi@th.btbu.edu.cn (Y. Chen).

https://doi.org/10.1016/j.jvlc.2018.02.003
Received 1 September 2017; Received in revised form 1 February 2018; Accepted 3 February 2018
Available online 01 June 2018
1045-926X/

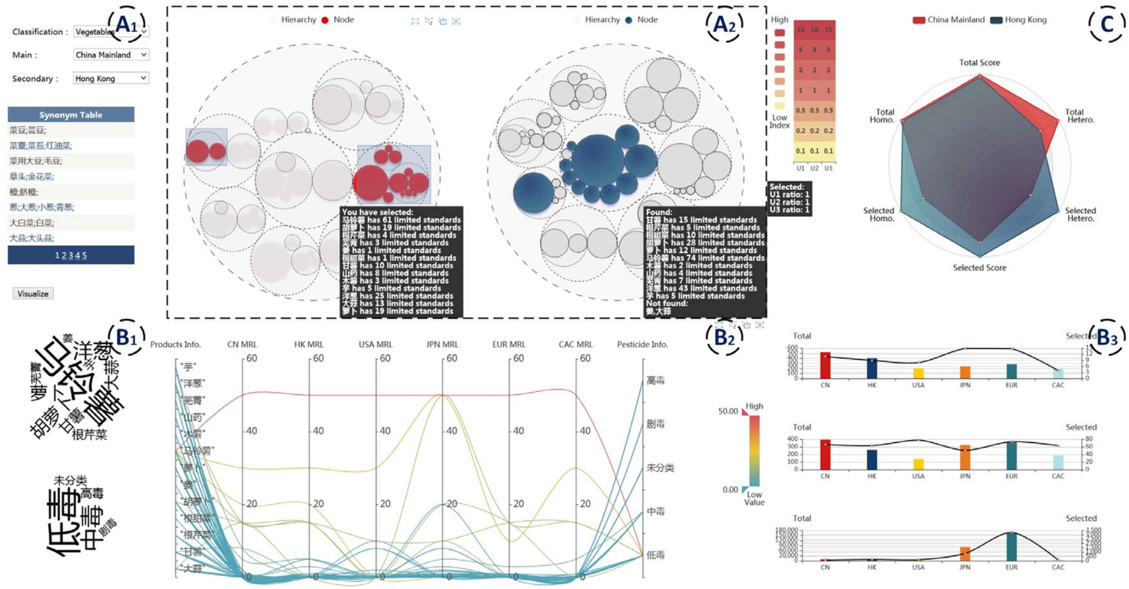

Fig. 1. A screenshot of the McVA system for multi-comparable visual analysis. The User Selected Panel (A1) for selecting and showing items from food categories, regions and synonym table. It is linked with Main Hierarchy Comparison View (A2) for visualizing the overviews for the hierarchies of two selected structures. Supplementary views include Bi-Attribute View (B1) for displaying the weights of two significant dimensions interactively, Multi-attributes Comparison view (B2) for comparing multidimensional values of different MRL standard data and Quick Fact View (B3) showing statistic of compared multi-region in the context; Evaluation View (C) for customizing view for multi-scale evaluation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Related work

2.1. Hierarchy comparison

In recent years, hierarchical data comparison related with people daily lives has been applied to many fields such as people behavior [1–3], public health [4–6], software and network analysis [7–10] and social entertainment analysis [11–13]. In visualization field, hierarchy comparisons were often classified into their structure and attribute comparison which can be abstracted as multiple trees or graphs that help users to discover the similar and different portions from diverse datasets [14].

No matter typical or abstract hierarchical data, for instance hierarchical topic modeling [15], comparison of these large-scale datasets that may contain thousands of points is often difficult due to their sheer scale involved. Three common methods exist for such comparisons, including juxtaposition, superposition, and animation [16]. Juxtaposition [17–19] involves the direct comparison of two datasets through combining the two trees to discover comparisons of the structures and then creating a new hybrid tree between them. Superposition [20–22] superimposes one dataset over another to highlight the differences between the two forms. Animation [23–25] uses encoded visualization to show the transition from one hierarchical data tree to another over time.

Among them, some of the approaches introduced explicit coding [26] or high-dimensional reduction method to quantify the comparison of multiple hierarchies. Bremm et al. [27] presented an effective approach using matrix and multiple views for comparing multiple trees task. Michael et al. [28] recently applied the idea of explicit encoding and juxtaposition in hierarchical phenotype patient data, namely PhenoBlocks, which was different from other patient-text visual analytics application. Other options exist to interpose alternative structure for the same trees [21] for those cases where the overall shape is uncertain or subject to change. Amenta et al. [29] and Hillis et al. [30] adopt MDS to calculate the distance between any two trees based on multi-dimensional properties and then map the results to a scatter plot. Recently, Liao et al. [31] purposed a cluster-based visual abstraction method to enhance the visualization of multivariate scatterplots in different levels which is received good performance. But these methods can only compare hierarchies in broad terms. Comparatively, a detailed comparison of leaf nodes in dynamic quantitative data sets [32, 33] is difficult without purpose-designed visualization.

So far only a few visual analytics solutions target the data analysis tasks in encoded scores calculation. TreeVersity2 [34] use color, direction and other code in nodes to display difference or percentage change. Tominski et al. [35] proposed to improve the color coding in the comparison task by combining some of the overlapping values. DAViewer [36] uses color mapping with the calculation to show the word segmentation results of different sub-structures.

2.2. Hierarchy comparison in food safety domain

The food safety problems are increasing attention in our daily life. Sometimes, it related to agricultural products, pesticides, additives, spatial geographic information with large scale and complex data characteristics.

The detection result data in food safety contains various agriculture products, pesticides and geographic information sampled by spots. After tested by food detection methods, domain experts use Chinese MRL standard to validate whether pesticides are abused or not. Since 2014, Chen et al. have proposed a number of visual analytics methods for comparing hierarchical data in pesticide residue detection results data to help experts to analyze these data. They firstly represented a hybrid layout algorithm using a double interrelated tree to display the relationship between regional hierarchies and pesticides classification [37]. Then they combined interrelated tree with sunburst to help user significantly explore the associations between different pesticides and products [38].

Considering multi-dimensions and geographic information visual analytics in detection result data, Chen and her fellows progressively purposed MCT [39], SunMap [40] and OSMT [41]. MCT modified multi-coordinates in treemap to compare detecting excessive pesticide residue distributed in each province of China. SunMap combined map, sunburst and matrix-heatmap to simulate spatial distribution of