McVA: A Multi-comparison Visual Analysis System for Maximum Residue Limit Standard in Food Safety

Yi Chen, Yu Dong, Yuehong Sun, and Jie Liang

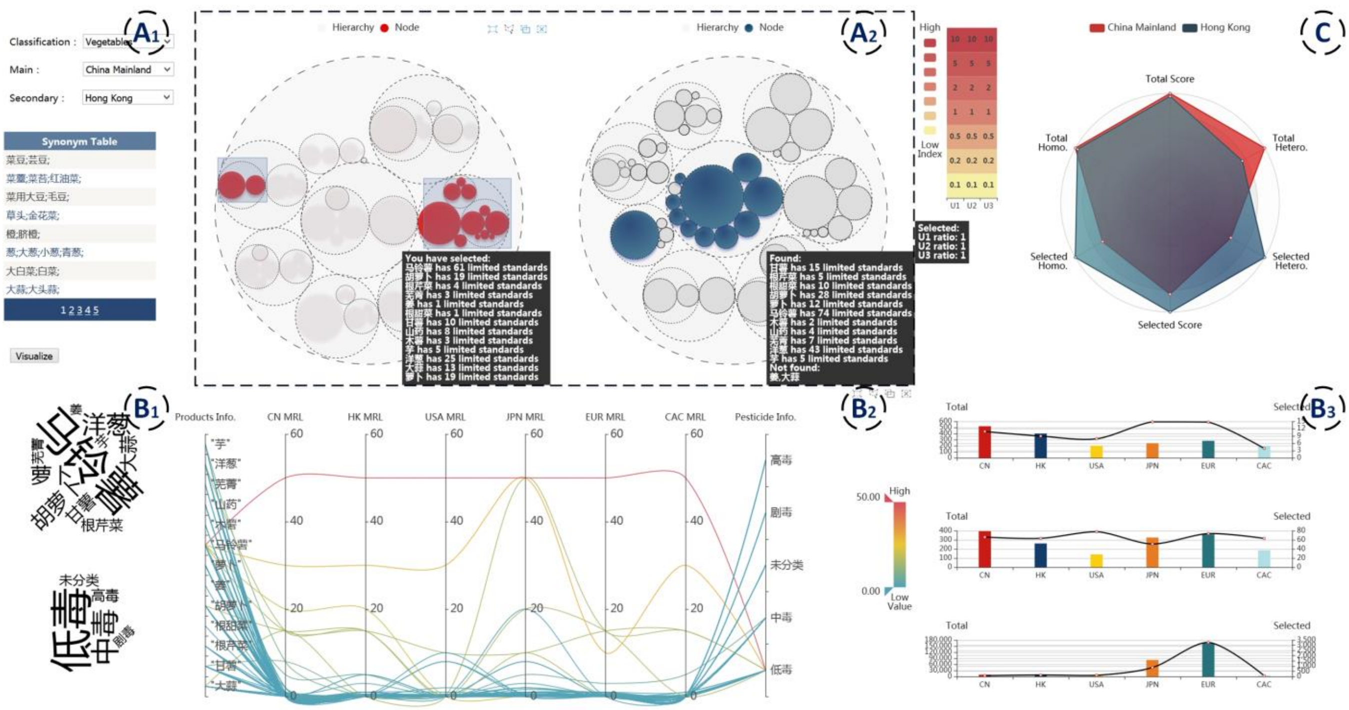

Fig. 1. Multi-comparison Visual Analysis System (McVA) Interface: (A_1) Selection Panel: selecting items to be compared, from the categories of food, regions and synonym table; (A_2) Main Hierarchy Comparison View: Visualizing the overviews for the hierarchies of two selected item; (B_1) Bi-Attribute View: displaying the weights of two significant dimensions selected in B_2 view; (B_2) Multi-attributes Comparison view: exploring multidimensional values of different food category;(B_3) Quick Fact View: showing statistic of compared region and other regions in the context; (C) Evaluation View: customizing view for multi-scale evaluation.

Abstract—Maximum residue limit (MRL) standard which specifies the highest level of every pesticide residue in different agricultural products plays a critical role in the food safety. However, such standards which relate to the classification of agricultural products and the characteristics of pesticides are complex and vary widely across different regions or countries. In this work, we present an interactive visual analytics system (McVA) to support multiple comparison and evaluation of MRL data standards. With a cooperative multi-view visual design, our proposed system links the hierarchies of MRL datasets and provides the capacity for comparison at different levels and dimensions. We also introduce a metric model for comprehensive evaluation of the completeness and strictness of an MRL, so that the MRL data with hierarchical structure and multidimensional attributes in the system can be visualized and compared more efficiently. We demonstrate the effectiveness of the approach by a case study of a real problem and positive feedback collected from domain experts.

Index Terms—Visual Analysis; Hierarchy Comparison; Multidimensional Data; Evaluation Metrics; MRL standard.

1 Introduction

Hierarchical data is one of the key research fields of information visualization and visual analysis. The analysis of hierarchical data has been applied to medical [1], biology [2], network analysis [3], software analysis [4] and other fields. In recent years, in-depth study of complex data sets found that hierarchical data is often accompanied by high-dimensional attributes. For example, the storage of files in an operating system: hierarchical storage of files can be represented using a tree structure, each file itself has multi-dimensional attributes [5] (file type, creators, document data and etc.).

Food safety, as a global problem, significantly affects the public health, especially in countries with large populations. A recent survey reported food safety was the most frightening issue in China. To improve food safety standards, domain experts often assess MRL across different food product, by comparing the variables in alignment with different regions or countries.

However, it is often difficult to make such assessment, due to the nature of MRL standards. MRL standard data are structured in multiple classifications of agricultural products, pesticides and other variables. Sometimes agricultural products vary in different region are named differently. Furthermore, each classification contains various attributes with multiple dimensions. Hence, comparing attributes in

- *Yi Chen, Yu Dong and Yuehong Sun are with Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. E-mail: {chenyi, sunyh}@th.btbu.edu.cn, 10011315258@stu.btbu.edu.cn.*
- *Jie Liang is with Faculty of Engineering and IT, School of Software, University of Technology Sydney, NSW, Australia. E-mail: jie.liang@uts.edu.au.*

such a multi-layered hierarchy can be extremely confusing. Last but not least, it is also a challenge to evaluate the MRL standards on a higher level.

In this paper, we present an interactive visual analytics system with a cooperative multi-view visual design, which links the hierarchies of MRL datasets and provides the capacity for comparison at different levels and dimensions. Main contributions of this work are list as below.

R.1 We implemented a visual analysis system called McVA for comparative analysis of multidimensional and hierarchical data. The system provides a suite of visualization techniques, which allows users to select, filter and interact with multidimensional attributes and explore the similarity and difference in the hierarchical structures of MRL standards.

R.2 We introduced a new evaluation model, based on the three dimensions of depth, breadth, and node data attribute values in a hierarchical structure. The user can adjust the index to evaluate and compare MRL datasets.

2 Related work

2.1 Hierarchy Comparison

Hierarchy comparison classified into structure and attribute comparison which can be abstracted as multiple trees or graphs that help users discover the similar and different parts of two datasets [6].

Comparison of large data sets that may contain hundreds of thousands of points is often difficult due to the sheer scale involved. Three common methods exist for such comparisons, including juxtaposition, superposition and animation [7]. Juxtaposition [8-10] 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 [11-13] superimposes one data set over another in order to highlight the differences between the two forms. Animation [14-16] uses visualization to show the transition from one hierarchical data tree to another over time.

Among them, some of achievements introduced explicit coding [17] and scores calculation to quantify the comparison of multiple hierarchies. S. Bremm et al [18] presented an approach using matrix and multiple views for comparing multiple trees task. TreeVersity2 [19] use colour, direction and other coding code in nodes to display difference or percentage change. Tominski et al. [20] proposed to improve the colour coding in the comparison task by combining some of the overlapping values. DAViewer [21] uses colour mapping with calculation to show the word segmentation results of different sub-structures. Others analyze two hierarchy by creating new visualizations to explore the differences. Amenta et al. [22] and Hillis et al. [23] adopt MDS to calculate the distance between any two trees based on multidimensional properties and then map the results to a scatter plot. But these method can only compare trees in broad terms. Other options exist to interpose alternate structure for the same trees [24] for those cases where the overall shape is uncertain or subject to change. Comparatively, detailed comparison of leaf nodes in dynamic quantitative data sets [25, 26] is difficult without purpose-designed visualization.

2.2 Hierarchy Comparison in Food Safety

The food safety problems are increasing attention in our daily life. It related to agricultural products, pesticides, additives, spatial geographic information and etc., with huge and complex data characteristics.

In recent years, Chen et al. have proposed a number of visual analysis methods for comparing hierarchical data in pesticide residue detection results. They represent a hybrid layout algorithm using double interrelated tree to display the relationship between regional hierarchies and pesticides classification [27]. Then they combined interrelated tree with sunburst to help user explore the associations between different pesticides and products [28].

Considering multi-dimensions and geographic information visual analysis in these data. Chen and other fellows progressively purposed MCT [29], SunMap [30] and OSMT [31]. 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 display spatial distribution of detection information interactively. OSMT used time-vary treemap layout algorithm to present pesticide residues detection results data and proposed a new metric (TVA) to evaluate the algorithms from new perspectives.

3 MRL in food safety

This section introduces the background on multi-region agricultural products, pesticide information and MRL standard data used. The analytical tasks will be discussed.

3.1 MRL Data

Each country or regional organization, respectively, has declared laws and regulations regarding the maximum residue limit (MRL) of pesticides and food additives. MRL standard refers to the maximum residue limit allowed by a pesticide or food additive in an agricultural product, the one which exceeds this value will be judged as an illegal product.

These limited standards can not only curb to a certain extent the occurrence of excessive pesticides or additives but also establish trade barriers between countries. Due to their later adoption, there was a gap between mainland China's MRL standards and those of other developed countries or regional organizations. It is a crucial to deeply study and compare this data, which can help domain experts to discover Chinese shortfalls and promote them to adopt international standards as quickly as possible.

In the past few years, we have been working closely with some experts in this field. We have implemented a pesticide detection information platform to manage information from collection, detection and then to data storage. By the end of 2016, we collected MRLs standard data from six organizations in countries/regions, including mainland China (CN), Hong Kong (HK), United States (USA), Japan (JPN), the European Union (EUR) and Codex Alimentarius Commission (CAC), with the results shown in Table 1:

Table 1. The amount of MRLs data in six countries or regions

	CN	HK	USA	JPN	EUR	CAC
MRLs	4140	7180	5657	44137	168382	3510
Products	330	316	389	254	368	315
Pesticides	433	358	313	592	588	154

In this paper, we used the multi-regional MRL standard data in fruits and vegetables for this analytic and comparative task.

3.2 MRL Characterization

Both the pesticides and agricultural products information, attached to multi-region MRLs standard, are complex datasets. Pesticide data contains multidimensional attributes such as pesticide toxicity, use, molecular formula, and isomers, etc. Each country or region has their own agricultural product classification. In other words, multi-region MRL standards are a kind of multidimensional and hierarchical data.

3.3 Related Analysis Task

As the country's agricultural products classification has high hierarchical structure characteristics, domain experts often compare the scale from multi-region classification.

According to the actual needs, we complete a list of analytical tasks.

R.1 Hierarchy comparison: We need to have an overview of the classification of multi-regions' agricultural products. And we also need to compare the classification of same agricultural products in different regions in detail.

Firstly, we need to compare the same agricultural products in different classifications from each multi-regional area and have a general impression of their characteristics

Secondly, we need to quickly identify the two regions of the classification of agricultural products in heterogeneous situation. That is, the agricultural products that are included in A, which are not included in B, or the agricultural products that are not included in A, but is contained in B.

Thirdly, we need to compare the amount of data provided by the same agricultural product in multi-regions. The more standard data for an agricultural product, the more stringent and complete we consider the data to be.

R.2 Nodes attribute comparison: Comparison of the number of different pesticides in same agricultural products and the MRL values in the heterogeneous structure, and the comparison of the number of heterogeneous agricultural products.

In comparing the leaf node attribute values, we mainly compare the MRL standard limits of the same pesticides under the same agricultural products in each region. The lower the threshold value, the more stringent the limited standard.

R.3 Hierarchy scorings: Scoring by multi-scale with the hierarchical structure and multidimensional attributes in two isomorphic hierarchical trees. Users can compare and explore the structural elements of their choice.

The values need to be quantified and contrasted in different hierarchical structures, both homogeneous and heterogeneous.

4 McVA Pipeline

McVA is a web-based application developed under the framework of ASP.net. We use Oracle 11g as database, and the visual analysis module is implemented using D3.js and echarts.js. Fig. 2 shows the McVA's system architecture and visual analytic pipeline.

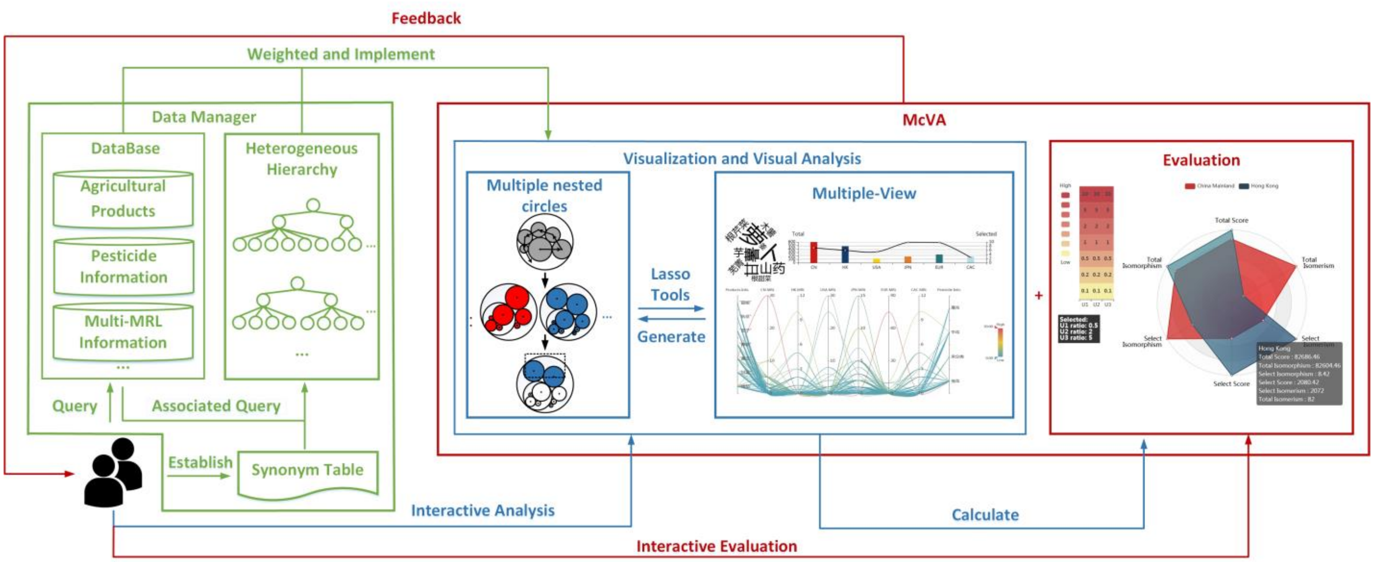

Fig. 2. McVA comprises two major components: visualization and visual analysis zone and evaluation zone. McVA needs to be supported by preprocessed data, together with synonym table established by the user named data manager. Multiple nested circles and other visualization figures constitute visualization and visual analysis zone. The evaluation zone generates visualization zone and rates from multi-scale.

5 McVA Approach

5.1 Synonym Table

In this paper, while the synonym data of agricultural products is provided by the relevant companies, we establish it by ourselves. When the user initiates the query request, the system queries the leaf nodes of the multiple heterogeneous structures with the synonym table and generates multiple heterogeneous hierarchies to reduce the loss of heterogeneous hierarchical data.

We set the names of agricultural products in mainland China as the primary key of synonyms associated with the names of other countries and regions, or the synonyms. As shown in Table 2 below.

Table 2. Synonym table example in Chinese and English

Product in Chinese classification	Slang in Chinese	Slang in English
芸薹	油菜, 油白菜, 苦菜	Cole, Rapeseed
芫荽	香菜	Caraway, Coriander
蘑菇	肉菌, 蘑菇菌	Mushroom, Agaricus
扁豆	鹊豆, 豆角, 白扁豆	Lentil, Hyacinth Bean

5.2 Data Abstract

Before the generation of multiple hierarchies, the synonyms are used to query the agricultural product classification data of each region. The processing as follows:

Step 1. We set the name of the Chinese agricultural products as the primary key, query the relevant classifications and agricultural products in other regions and establish the related relationships. The specific agricultural product classification and name shall prevail in each country.

Step 2. By the classification of different agricultural products in various regions, we extract agricultural product information from heterogeneous trees. As shown in Figure 3, the tree T_1 uses the synonym table to associate the query and abstract the structures of tree T_2 and tree T_3. The leaf nodes represent various agricultural products, and the nodes present different agricultural products. We can see that the distribution of N_1-N_5 in T_2 and T_3 may be different from the location of nodes in T_1.

Step 3. The association of agricultural products and the MRL standard in each region is used to collect the number of MRL standards for all pesticides in one certain agricultural product as the weight of the leaf node of the agricultural product. We expand the product leaf node as a root to generate a new subtree with multidimensional data and generate the weighted tree structure of product classification in multi-regions. As shown in Fig. 4, the tree T_1 in this figure is used to generate a subtree with depth of 1, under the leaf node of each agricultural product after querying the MRL standard data. The new subtree leaf nodes are a variety of MRL value for all pesticides in one individual agricultural product. We note the number of pesticides under each node and record as the weight of the agricultural node, gradually form multiple weighted trees.

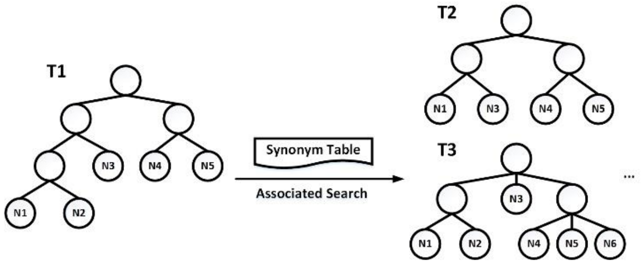

Fig. 3. The process of associated searching and generating multiple heterogeneous structures.

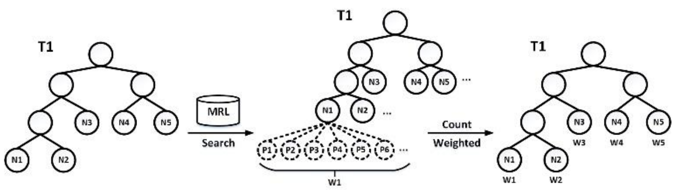

Fig. 4. The process of generating weighted hierarchical data.

5.3 Nested circle with lasso tool

Using algorithms to generate nested circles for visual expression, programming to implement the lasso tool helping user to interact with the data of interest. The process is as follows:

Step 1. The agricultural products are sorted by their weight in descending order. We use nested circle algorithm to generate multi-regional agricultural product classification with the formation process shown in Figure 5 (a). The nodes are arranged in a spiral from largest to smallest weight. After the arrangement is completed, the distance between the two nearest circumscribed points is set as a new diameter to draw the parent nested circle. The weight of the new circle is the sum of its sub-circle weights. The process is repeated recursively step by step, until the root circle is completed.

Step 2. Using JavaScript, we have programmed the lasso tool visual analysis method. This allows the user to use the matrix and custom shape tools to keep multiple choices to circle the product data in the main graph, allowing the associated agricultural products in the secondary graph will be highlighted. The way to determine whether the node is selected is the circle area passes through the center.

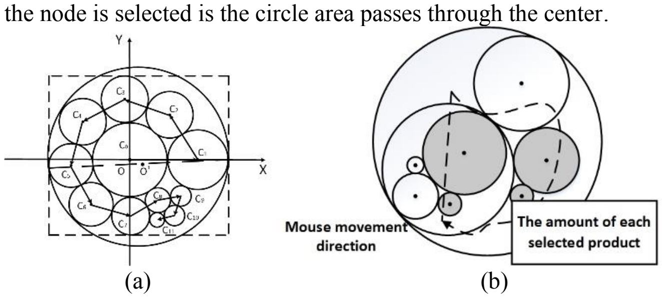

Fig. 5. Nested circle. (a) The process of sorting leaf nodes and creating the father node. (b) Using lasso tool to select the leaf nodes.

5.4 Evaluation Metric Establishment

We introduced three joint indices in one metric to compare two hierarchical structures. Due to the particularity of the MRL standard data for pesticide residues, the MRL data values under the same kind of agricultural products and pesticides, are compared. The smaller the value, the more stringent it is considered to be.

To further describe the calculation process, we define the following variables, as shown in Table 3:

Table 3. Mathematical notations and descriptions

Notation	Description
t	The number of the tree structure.
k	The level of the tree.
a	The name of agriculture product name or product classification.
a_t^k	The agriculture product of product classification node in level k tree t
V_t^i	One of the element's value under leaf node i in tree t
$C\left(a_t^k\right)$	The set of children nodes element under node a in level k tree t
$P\left(a_t^k\right)$	The set of $\left\{C\left(a_t^k\right)\cap C\left(a_{t'}^k\right)\right\}$ the element.

The equation of the joint indices is accomplished by our visual analysis and evaluation approach. As Fig. 3 shows in section 5.2, in multiple hierarchies, there are leaf nodes with the same data name, we called them the homogeneous data. And their own unique nodes called the heterogeneous data. By the combination of these two parts, the weight of each leaf node is collectively calculated. Fig. 5 shows the same composition and the unique data between the two different hierarchical structures which will be calculated separately.

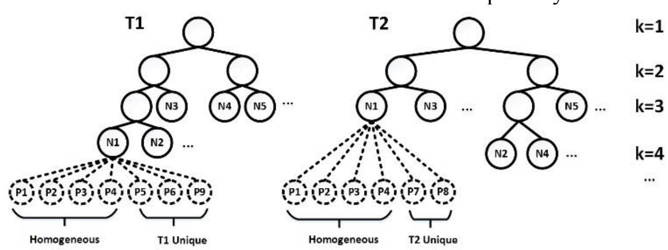

Fig. 5. Two different hierarchies have the homogeneous and their own unique (heterogeneous) data.

5.4.1 Computing the homogeneous data

Due to the specificity of the MRL standard data, the lower a stated value, the higher the calculated result should be. We need to design a decreasing function to describe it. However, the general subtraction function is ineffective in representing a large magnitude difference, so we develop a dynamically changing logarithmic function to quantify the same composition data between the two hierarchies. The detailed steps are as follows:

Step 1. Construct the gain rate: We define a variable that is used to describe the larger value in the same node between two hierarchical structures t and t':

$$V_{\max}^{i} = \left\{V_t^i,\ V_{t'}^i\right\}_{\max} \tag{1}$$

After that, we use a rate α to describe the gain between 0 and 1:

$$\alpha = \frac{\left|V_t^i - V_{t'}^i\right|}{V_{\max}^i} \tag{2}$$

Step 2. Construct logarithm: To reduce the magnitude gap in different data, we use a dynamic logarithmic function to describe it. With the logarithm we choose a variable β between 0 and 1 which is the current value divided by the sum of the current name values in both structures:

$$\beta = \frac{V_t^i}{V_t^i + V_{t'}^i} \tag{3}$$

Step 3. Computing the homogeneous data: We use logarithmic sum to represent all of the same composition data:

$$Na_t^k = \sum_{i \in P\left(a_t^k\right)} \log_\alpha \beta \qquad (4)$$

5.4.2 Computing the heterogeneous data

The heterogeneous data is meant to only exist in one hierarchy. So we use a cardinality set to describe it. We add the number of unique data under node *a* as the result. Finally, all of the unique data in t is summed and computed as follows:

$$Ma_t^k = \left| C\left(a_t^k\right) - P\left(a_t^k\right) \right| \qquad (5)$$

5.4.3 Computing leaf node weight

The weight of one leaf node should be equal to the addition of the same composition and its own unique composition. However, we consider that the same composition is not in the same magnitude level with the different composition, we add 2 multipliers. U_1 describes the multiplier of the same composition and U_2 describes the multiplier of the different composition. User can adjust it according to actual needs.

The weight of the leaf node on last level k_{max} should be:

$$Wa_t^{k_{max}} = Na_t^{k_{max}} \times U_1 + Ma_t^{k_{max}} \times U_2 \qquad (6)$$

5.4.4 Computing node weight

We extend the definition of node weight to the common level *k* and add U_3 to describe the different level multiplier. So, any node weight can be defined as:

$$Wa_t^k = \begin{cases} \sum_{i \in C\left(a_t^k\right)} Wa_t^{k+1} \times U_3 & k < k_{max} \\ Wa_t^{k_{max}} & k = k_{max} \end{cases} \qquad (7)$$

Eq. 7 provides a computing method to calculate the weight of the root node recursively. We can adjust the multiplier and use it to compare multiple hierarchical structures.

In this section, the node weight in any level can be extended to the other two indices. One is all the homogeneous data we use to compare the summary of attribute values; the other one is the number of all the heterogeneous parts of the statistics.

6 McVA System Design

In this section, we describe a system of visual analytics named McVA that assists users in generating solutions for exploring and comparing hierarchical data. It contains two components as follows:

6.1 Interface for Hierarchy and Dimension Comparison

Due to the computational power of modern machines, users can display the hierarchical structure and attribute values of each leaf node. In this paper, we use multiple cooperative views to help users combine this feature with their needs and analyze them.

6.1.1 Hierarchy selection panel

From the top to bottom of the user selection zone (in Fig. 1 A_1), the user can select the root of product classification, the main and secondary view in the next part of a nested circle. Under these, there is a synonym table which is established by the user displaying the same agricultural product from multi-regions. Users can click "Visualize" button to draw two nested circles in next zone.

6.1.2 Main hierarchy comparison view

The main hierarchy comparison view should clearly represent the level and other products hierarchical information. Traditional treemap or matrix-based visual methods can make good use of space, but it may not clearly display the unique information in different levels. Based on the actual needs, nested circle is selected as the main view.

Main hierarchy comparison view (in Fig 1 A_2) adopts nested circular treemaps, and shows the classification of agricultural products in two countries or regions selected by the user. It is designed to display the classification information as a dotted circle and distinguish agricultural products in different colored leaf nodes.

The process of formation is described in section 5.3. In connection with MRL database, each circle's colored node indicates the amount of all MRL standard data involved in this product. Users can select the node's data via lasso tool in the main view, and the secondary view highlights the corresponded node immediately. The lower right corner of the view will show the amount of selected agricultural products in the MRL standard data. It helps the user not only discover the unique agricultural products in two circles quickly but also to compare the amount of MRL data in the same leaf node product.

The lasso tool, as an effective interaction method, makes the nested circle as an originator of the visual analysis task. This allows the user to explore the generated data through the word cloud, parallel coordinates, line-bar charts and evaluation zone with other related information.

6.1.3 Bi-dimension view

We use two word cloud graphs as bi-dimension view (in Fig 1 B_1) to reveal more agricultural product and pesticide information.

The top word cloud, called the product word cloud, shows the product selected by users. The font size represents the amount of MRL standard data that each product contains, the same as the tooltip in the main nested circle view.

The bottom word cloud, called the toxicity word cloud, shows the toxicity of the pesticide selected by users. The font size represents the amount of pesticide involved in each toxicity.

Both word clouds can interact through parallel coordinates. The user also can click on any word cloud in parallel coordinates to display the relevant agricultural products and toxic information, so as to carry out the following analysis of MRL value in multi-regions.

6.1.4 Multi-dimension view

Multi-dimension view (in Fig 1 B_2) shows the distribution of MRL standard data in each region for the agricultural products selected by the lasso tool. It presents 8 dimensions as pesticide information, mainland China MRL, Hong Kong MRL, USA MRL, Japan MRL, the Europe Union MRL, CAC MRL and pesticide toxicity.

As the following Fig.6, each of the products contains multiple incidences of pesticide and MRL standard data. To optimize the display, we set the last row of parallel coordinates as the pesticide toxicity column.

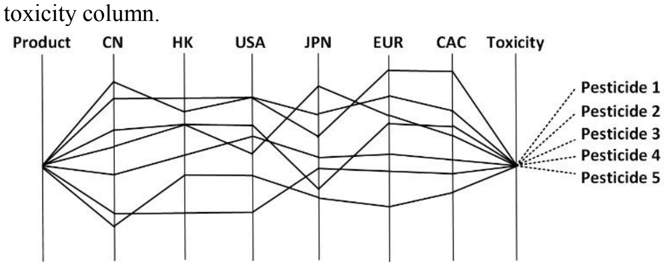

Fig. 6. The toxicity column contains the description of the pesticide in parallel coordinates

We focus on comparing and studying the difference between mainland China's MRL standard data and other regions. So use color to identify the value of MRL in mainland China. When the color becomes red on for the standard value, it means that these values are not stringent and more dangerous.

6.1.5 Quick fact view

We put three line-bar charts to help the user comparing other information horizontally and vertically (in Fig. 1 B_3). From top to bottom, each of the charts shows products, pesticide information, and MRL data. We use two Y-axes to display data; all the total data coordinates are on the left, selected data is in the opposite direction. All the three charts can not only compare the same classification data but explore the similarity between selected and total data.

6.2 Evaluation Interface

The evaluation index established in Section 5.4 can be used to compare and evaluate two hierarchical structures from six scales.

When considering the interaction, we compare two structures from the perspective of the overall structure and the selected substructures by users. Each of them contains three evaluated indices.

6.2.1 Overall structure

The overall structure contains Total Score, Total Homo and Total Hetero three comparison scores portion. They can be compared MRL data under two agricultural products' trees from macro perspective.

R.1 Total Score: The total score refers to the sum of the weights of all child nodes under the root node, like the calculation in Eq. 7. It can give the user an overview and contrast to the overall structure.

R.2 Total Homo: Total Homo is a comparison score of the leaf node values of the homogeneous portion in the two trees. It can quantize and compare the value of the same composition data.

R.3 Total Hetero: Comparing with two trees, the number of all heterogeneous data in tree *t* called isomerism in total structure *t* score. It can be used to create quick statistics from the heterogeneous data.

6.2.2 Selected substructures

The selected substructures' scores can be changed by user selected dynamically to show these three aspects.

R.1 Selected score: The selected substructure score is the sum of the weights of all nodes selected by the user. It can be through the user interaction with the behavior analysis, selected substructure scores can be calculated dynamically.

R.2 Selected Homo: Selected Homo is a comparison score of the selected leaf node values of the homogeneous portion in the two trees. Users can quickly compare the MRL values in the same agricultural products or the same product classification.

R.3 Selected Hetero: Comparing with the nodes selected by users, the number of unique data indices is called Selected Hetero. The user can quickly compare the number of unique data indices under a particular product or classification.

6.2.3 Visual evaluation

Considering the actual needs of the data, we calculate and display the node scores, the homogeneous part, the heterogeneous part and its hierarchical information, which may account for different proportions. We use U_1, U_2, U_3 in Eq.7 and add an operable heat map to help the user adjust the multiplier. The initial value of each multiplier is 1, and the value range is from 0.1 to 10.

The radar is connected to the heat map. From the top of the radar, going counterclockwise, the axes are Total Score, Total Homo, Selected Homo, Selected Score, Selected Hetero and Total Hetero.

7 CASE STUDY

We conducted a case study with a specific example and system verification for actual analytics task and also commissioned a domain expert review to demonstrate the effectiveness of McVA.

7.1 Specific example

We have conducted a practical task about comparing rhizome and partial bulb vegetable agricultural products in mainland China and Hong Kong.

As shown in Fig. 1, the user selected the vegetable classification, mainland China as the main view and Hong Kong as the secondary. After user clicked the "visualize" button, the nested circle showed the distribution of agricultural products in these two regions. The user can use the lasso tool to circle the data of interest. When he selects all the rhizome and partial bulb products in mainland China (dashed line), the Hong Kong view highlights the corresponding agricultural products. Comparing these two nested circles, he can clearly see the classification of agricultural products at all levels. As shown in Fig. 7, the classification nodes were in black, the common products of two regions were in white and the unique products were in grey nodes.

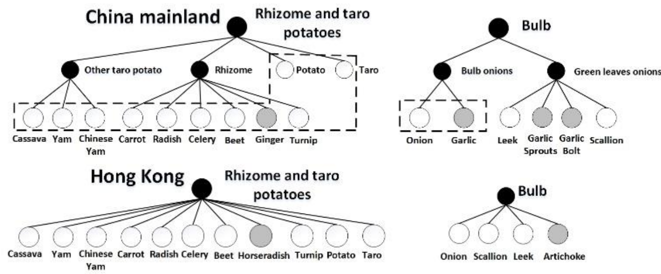

Fig. 7. Different classifications of rhizome and partial bulb vegetables in China mainland and Hong Kong

The value of each product node (MRL data) is shown in Table 4:

Table 4. The value of selected agricultural product node in China mainland and comparing Hong Kong in highlight

Product name	Value in China mainland	Value in Hong Kong
Potato	61	74
Onion	25	43
Radish	19	12
Carrot	19	28
Garlic	13	0
Yam	10	15
Chinese Yam	8	4
Taro	5	5
Celery	4	5
Cassava	3	2
Tumip	3	7
Beet	1	10
Ginger	1	0

Note: the node value represents how many MRL standards is under this agriculture product.

By comparing the above table and graph, we can see the difference between agriculture product classification and the value of the MRL data under each selected agricultural product.

Under the rhizome and taro potato classification, there are two sub-categories in mainland China, which divides these agricultural products into rhizome and other sub-classifications. Hong Kong only has one layer of classification. We can see that mainland China under the rhizome classification is more complete than Hong Kong.

The selection of onions and garlic under the China mainland classification is also in the second layer of the bulb, while onion in Hong Kong is located in the first layer and does not contain garlic data.

Comparing the value of each node selected by the user, we can see the amount of MRL data in each product. Although both mainland China and Hong Kong have their unique MRL data like ginger, horseradish, etc. But overall, the number of homogenous products in Hong Kong's MRL data is much larger than in mainland China.

Using the word cloud and parallel coordinates can effectively show the MRL value of user-specified agricultural products and the distribution of pesticide toxicity. Using line-bar charts to interact with them can compare the supplement information and find homogeneous and heterogeneous data quickly.

In general, the user always prefers to focus on agricultural MRL data in higher toxicity. Higher toxicity pesticides stipulated in MRL data reflects the standard stringency in multi-regions. With the selected high toxicity pesticide, Table 5 shows the number of mainland China agricultural products and MRL standard are ahead of others, the provisions of pesticide slightly behind Hong Kong, Japan, and the European Union.

Table 5. The amount of selected MRLs data in six countries or regions

	CN	HK	USA	JPN	EUR	CAC
Products	5	5	4	4	4	4
Pesticides	7	8	6	8	9	7
MRLs	10	5	2	6	7	3

Two hierarchies can be quantified by three evaluation indices that user can compare them interactively. McVA scores the whole and the two selected structures by six scales. When the radar map is automatically generated, U_1, U_2, U_3 are set to 1. The user selected all the classifications in rhizome products, the detailed result as follows:

Table 6. The initialized result of all vegetables and rhizome classification in China mainland and Hong Kong

	TS	THo	SHo	SS	SHe	THe
CN	39870.18	39647.15	2.62	136.62	105	852
HK	38762.15	38721.15	3.67	165.67	194	650

Note: TS, THo, SHo, SS, SHe, THe represent Total Score, Total Homo, Select Homo, Select Score, Select Hetero, Total Hetero.

Users can interact with operable heat map to adjust the size of the multiplier from 0.1 to 10 to calculate the customized scores.

If users believe that the isomorphic portion and the data level are more important, they can increase these multipliers to get the new scores. They adjusts U1 to 2 and the new results as following:

Table 7. The result of all vegetables and rhizome classification in China mainland and Hong Kong with U_1=2

	TS	THo	SHo	SS	SHe	THe
CN	135825.89	135602.89	11.59	145.59	105	852
HK	115488.25	115447.22	7.34	169.34	194	650

And then, he adjusts U_3 to 10 to increase the level of the proportion of these scores, the results following:

Table 8. The result of all vegetables and rhizome classification in China mainland and Hong Kong with U_1=2 and U_3=10

	TS	THo	SHo	SS	SHe	THe
CN	1356251.89	1356028.89	115.89	249.89	105	852
HK	1154513.25	1154472.25	73.43	235.43	194	650

Finally, he changes U_1, U_3 to the initial value and adjusts U_2 to 2 to increase the heterogeneous scores, the results following:

Table 9. The result of all vegetables and rhizome classification in China mainland and Hong Kong with U_2=2

	TS	THo	SHo	SS	SHe	THe
CN	40801.18	39647.18	2.62	982.62	330	3644
HK	38811.15	38721.15	3.67	651.67	297	2032

The initialized result represents all the agricultural products at the first level, and the homogeneous and heterogeneous levels are in the same proportion. We can see that the results in mainland China are worse than those in Hong Kong. The results of different multipliers are also different. There may be an order of magnitude difference between the elements within. By comparing the other results we can see the expansion of U_1, U_2 and U_3 may affect the comparison scores of the two trees. All the multiplier settings need to follow the actual needs of the operation.

7.2 System feasibility verification

We further performed another user study to validate the system feasibility with 50 students from four majors in School of Computer Science and Technology, who were unfamiliar with visualization but familiar with the use of SQL and database tools.

Then we divided them into two groups. After trained, students in first group started to use McVA to analyze real visual comparison work. The second group students used SQL to search the result. During the whole experiment, each participant was asked to perform a task, answering the following questions in Table 10.

Table 10. Questions designed on MRL data

No.	Questions
Q1	How many kinds of fruits are classified in 6 countries or regions?
Q2	What kinds of vegetables are there under Chinese solanaceous?
Q3	Please filter all the MRL standard data for cucumber in high toxicity in 6 countries or regions.
Q4	Please find out the quantity of agricultural products, pesticides and MRL standard data with brassica classification in multi-regions.
Q5	Try to evaluate MRL data of tomato or solanaceous vegetables in China mainland and HK.

Note: the MRL data evaluation was considered from two aspects: hierarchy and MRL standard values. Participants would be asked questions like: Which country in the tomato MRL data is more stringent? Which country is more complete in the classification of solanaceous vegetables?

The average accuracy of the answers and the average time the users spent during the experiment were the two aspects we recorded which were compared in Fig. 8 as below.

From the results, we can see that the time cost of using McVA is relatively much shorter than using database tools and the accuracy of using system works better at the same time.

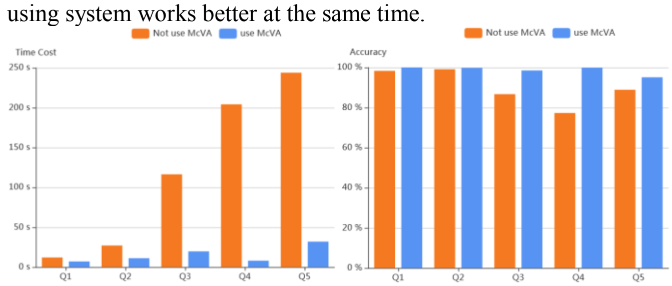

Fig. 8. The time cost and accuracy results recorded in two histograms

7.3 Domain expert interview

We analyzed some specific examples and collected feedbacks from the domain experts by conducting one-on-one interviews. The representative feedbacks were summarized as follows:

Visual Design and Interactions. All experts confirmed that our McVA system is well designed and user-friendly. In particular, our lasso tools interactive design and evaluation zone received high praise from the domain experts. They believed that the system could be easily used and evaluated. Expert A who is active in visualization field commented, “It is useful to give consideration to displaying different hierarchical information and nodes of multidimensional data. The interaction is flexible and can assist users to explore and compare the data quickly".

Usability and Improvements. The experts appreciated our system has a great significance. They all agreed that McVA is useful in not only comparing the MRL data in multi-regions but also applying to other comparative hierarchical multidimensional data analysis tasks. Apart from those above, experts also provided some valuable suggestions. Expert B with a strong agriculture experience mentioned, "When I compare the structure of two hierarchical datasets, it would be nice if I could show the substructure data I selected in the same visualization graph rather than rely on my eyes to pick it out."

8 Conclusion and Future Work

In this paper, we thoroughly study the problem of comparing and analyzing hierarchical structures, including the hierarchical information and multidimensional node attributes. We implement a visual analytics system called McVA using multiple views and scales to help the user explore and evaluate the MRL and other related data in multi-regions. We also conduct a series of user case to demonstrate the effectiveness of McVA.

Our contributions are summarized below:

Firstly, McVA can show the classification of agricultural products in different regions. The user can quickly determine the level of agricultural products and node (data) size.

Secondly, the user can use different interaction methods to compare the homogeneous and heterogeneous products in multi-regions effectively.

Thirdly, McVA supports other useful information on agricultural products, pesticides and MRL standards through a multi-view visualization.

Fourthly, considering the level of the structure, the homogeneous node values and the number of heterogeneous nodes, we propose a new model for calculating the evaluation of hierarchical data.

Although our system is primarily designed for multi-regional MRL data, it can be slightly changed in the assessment formula to adapt to other multidimensional and hierarchical data analysis tasks easily.

While our approach is effective, there are a few aspects to be improved further. In future work, we will improve the visualization to accommodate hierarchical data from different structures in the same graph that shows the hierarchical information intuitively.

Acknowledgments

This work is supported by "Twelfth Five-Year" National Science and Technology Program (No.2012BAD29B01-2), National Program on Basic Science and Technology Project (No.2015FY111200) and Beijing Science and Technology Program (No.Z151100001615041).

References

[1] Girardi, Dominic, et al. Using Concept Hierarchies to Improve Calculation of Patient Similarity. Journal of Biomedical Informatics, 2016,63:66.

[2] Hu, Yifan, and Y. Koren. Extending the spring-electrical model to overcome warping effects. Visualization Symposium. IEEE Pacific Visualization Symposium, 2009:129-136. [doi: 10.1109/ PACIFICVIS. 2009.4906847]

[3] Shi, Lei, et al. Hierarchical Focus+Context Heterogeneous Network Visualization. Visualization Symposium IEEE, 2014:89-96. [doi: 10.1109/PacificVis.2014.44]

[4] Beck F, Diehl S. Visual comparison of software architectures. Information Visualization, 2013,12(2):178−199. [doi: 10.1177/14738 71612455983]

[5] Teoh S T, Ma K L. RINGS: A Technique for Visualizing Large Hierarchies. Revised Papers From the, International Symposium on Graph Drawing, 2002:268-275. [doi: 10.1007/3-540-36151-0_25]

[6] Graham M, Kennedy J. A survey of multiple tree visualization. Information Visualization, 2010,9(4):235−252. [doi: 10.1057/ ivs.2009. 29]

[7] Li YL, Li GQ, Dong XJ. Survey on visualization of tree comparison. J. Software, 2016,27(5):1074−1090. [doi: 10.13328/j.cnki.jos.004957]

[8] Munzner T, Guimbretière F, Tasiran S, Zhang L, Zhou YH. Treejuxtaposer: Scalable tree comparison using focus+context with guaranteed visibility. ACM Trans on Graphics, 2003,22(3):453−462. [doi: 10.1145/882262.882291]

[9] Telea A, Auber D. Code flows: Visualizing structural evolution of source code. Computer Graphics Forum, 2008,27(3):831−838. [doi: 10.1111/ j.1467-8659.2008.01214.x]

[10] Stewart CA, Hart D, Berry DK, Olsen GJ, Wernert EA, Fischer W. Parallel implementation and performance of FastDNAml: A program for maximum likelihood phylogenetic inference. In: Proc. of the ACM/IEEE Conf. on Supercomputing. 2001:22−32. [doi: 10.1145/ 582034.582054]

[11] Morse D R, Ytow N, Mcl D, et al. Comparison of multiple taxonomic hierarchies using TaxoNote. IEEE Symposium on Information Visualisation, 2003, 24(1):16-30.

[12] Lee B, Robertson GG, Czerwinski M. CandidTree: Visualizing structural uncertainty in similar hierarchies. Information Visualization, 2007,6(10): 233−247. [doi: 10.1007/978-3-540-74800-7_20]

[13] Tu Y, Shen HW. Visualizing changes of hierarchical data using treemaps. IEEE Trans. on Visualization and Computer Graphics, 2007,13(6): 1286−1293. [doi: 10.1109/TVCG.2007.70529]

[14] Robertson G, Cameron K, Czerwinski M, Robbins D. Animated visualization of multiple intersecting hierarchies. Information Visualization, 2002,1(1):50−65. [doi: 10.1057/palgrave.ivs.9500002]

[15] Card SK, Nation D. Degree-of-Interest trees: A component of an attention-reactive user interface. In: Marsico MD, ed. Proc. of the Working Conf. on Advanced Visual Interfaces. Gallipoli: ACM Press, 2002. 231−245. [doi: 10.1145/1556262.1556300]

[16] Heer J, Card SK. DOITrees revisited: Scalable, space-constrained visualization of hierarchical data. In: Proc. of the Working Conf. on Advanced Visual Interfaces. Gallipoli: ACM Press, 2004. 421−424. [doi: 10.1145/989863.989941]

[17] Gleicher M, Albers D, Walker R, Jusufi I, Hansen C, Roberts J. Visual comparison for information visualization. Information Visualization, 2011,10(4):289−309. [doi: 10.1177/1473871611416549]

[18] Bremm S, Landesberger T V, Heß M, et al. Interactive visual comparison of multiple trees. IEEE Visual Analytics Science and Technology press, 2011:31-40. [doi: 10.1109/VAST.2011.6102439]

[19] Guerra-Gomez JA, Pack ML. Visualizing change over time using dynamic hierarchies: TreeVersity2 and the StemView. IEEE Trans. on Visualization and Computer Graphics, 2013,19(12):2566−2575. [doi: 10.1109/TVCG.2013.231]

[20] Tominski C, Fuchs G, Schumann H. Task-Driven color coding. In: Banissi E, ed. Proc. of the IEEE Int'l Conf. on Information Visualization. London: IEEE Computer Society Press, 2008. 373−380. [doi: 10.1109/IV.2008.24]

[21] Zhao J, Chevalier F, Collins C, Balakrishnan R. Facilitating discourse analysis with interactive visualization. IEEE Trans. on Visualization and Computer Graphics, 2012,18(12):2639−2648. [doi: 10.1109/ TVCG. 2012.226]

[22] Amenta N, Klingner J. Case study: Visualizing sets of evolutionary trees. In: Chuang WP, ed. Proc. of the IEEE Symp. On Information Visualization. Boston: IEEE Computer Society Press, 2002. 71−74. [doi: 10.1109/INFVIS.2002.1173150]

[23] Hillis DM, Heath TA, John KS. Analysis and visualization of tree space. Systematic Biology, 2005,54(3):471−482. [doi: 10.1080/ 106351505909 46961]

[24] Lee B, Robertson G G, Czerwinski M, et al. CandidTree: Visualizing Structural Uncertainty in Similar Hierarchies. Information Visualization, 2007, 6(3):233-246. [doi: 10.1007/978-3-540-74800-7_20]

[25] Burch M, Weiskopf D. Visualizing Dynamic Quantitative Data in Hierarchies - TimeEdgeTrees: Attaching Dynamic Weights to Tree Edges. Imagapp & Ivapp 2011 - Proceedings of the International Conference on Imaging Theory and Applications and International Conference on Information Visualization Theory and Applications, Vilamoura, Algarve, Portugal, March DBLP, 2011:177-186.

[26] Finis J, Brunel R, Kemper A, et al. Indexing highly dynamic hierarchical data. Proceedings of the Vldb Endowment, 2015, 8(10):986-997. [doi: 10.14778/2794367.2794369]

[27] Yi C, Zhang X Y, Chen H Q, et al. Hybrid Layout Algorithm for Double Interrelated Tree. Journal of System Simulation, 2014.

[28] Y Chen, YG Zhen, HY Hu, J Liang, KL Ma. Visualization technique for multi-attribute in hierarchical structure. J. Software, 2016(5):1091-1102. [doi: 10.13328/j.cnki.jos.004956]

[29] Chen Y, Zhang X, Feng Y, et al. Sunburst with ordered nodes based on hierarchical clustering: a visual analyzing method for associated hierarchical pesticide residue data. Journal of Visualization, 2015, 18(2):237-254. [doi: 10.1007/s12650-014-0269-3]

[30] Y Chen, X Lin, Y Zhao, Y Sun, X Zhang. SunMap: an associated hierarchical data visualization method based on heatmap and sunburst. Journal of computer aided design and computer graphics, 2014, 26(9):2160-2165. [doi: 10.3969/j.issn.1003-9775.2016.07.006]

[31] Chen Y, Du X, Yuan X. Ordered small multiple treemaps for visualizing time-varying hierarchical pesticide residue data. Vis Comput, 2017, 33:1073–1084. [doi:10.1007/s00371-017-1373-x]