TreeVersity: Comparing Tree Structures by Topology and Node Attribute Values

John Alexis Guerra Gómez, Catherine Plaisant and Ben Shneiderman
Department of Computer Science
Contact: {jguerrag, plaisant, ben}@cs.umd.edu
www.cs.umd.edu/hcil/treeversity

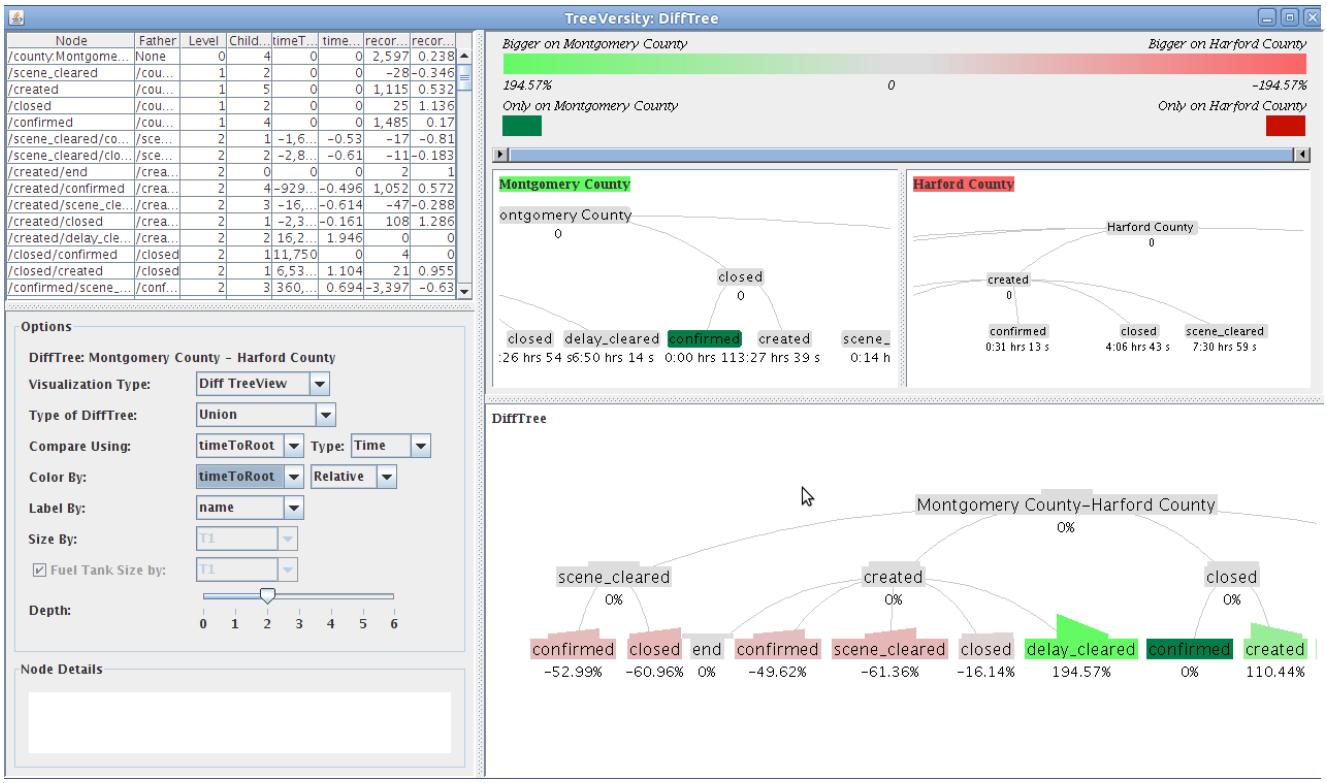

Figure 1 – Comparing the tree of possible sequences of actions performed by the first responders when attending traffic incidents in two different counties using TreeVersity. On the left part of the graphic is the control panel to manipulate the parameters, filter the different nodes and consult their details. On the right hand size are, on the top the two trees compared (Montogmery County vs Harford County) while on the bottom is displayed the DiffTree. From it, both topological (dark green and red) and node attributes differences (shades of light green to light red) can be identified.

Viewing individual tree structures has a rich history. But comparing two trees is a substantially greater challenge. Smartmoney's MarketMap [1] shows the change in price by color coding, and TreeJuxtaposer [2] shows the change in topology by highlighting matching nodes in two trees. These solutions, although effective for specific tasks, need enhancements to generalize. With this project we propose to deal with richer cases such as:

1. Positive and negative changes in *leaf* node's attribute values, *without* changes in topology. For example comparing the stock market's closing prices of two different days, assuming no stocks are created or deleted.
2. Positive and negative changes in *leaf* node's attribute values, *with* changes in topology. For example comparing the Federal Budget in two different years, for the different agencies. Since from a year to another there might be some dependencies that disappear or are created there might be some topological changes.

3. Positive and negative changes in leaves and *interior* node's attribute values, without changes in topology. For example comparing the statistics of the NBA players during the playoffs versus the regular season (Assuming that no changes in the team's payroll is made).
4. Positive and negative changes in leaves and *interior* node's attribute values, *with* changes in topology. For instance comparing the number of page visits for two months on a website hierarchy. Where some pages might be created or removed, and each page has a independent number of visits.

TreeVersity

TreeVersity is an interactive visualization tool design to help comparing tree structures both in topology as in node attribute values. The tool is been designed to support the following use cases:

1. Given two trees, compare them and discover by user exploration and interaction, what is different and what is similar between them.
2. Same as the previous case, but provide a ranking of the most significant differences and similarities, and guide the user through them.
3. Given a forest (a group of trees) and one selected tree as a base, find the most similar (and the most dissimilar) hierarchies, and display their similarities (and differences).
4. Given a forest, create an average tree that represents the most common characteristics among the trees.
5. Find Given a forest, find the most peculiar trees. That is trees that are the most different to the average.

Currently, we are working on implementing and evaluating the first use case. A screen capture of the current prototype can be seen on Figure 1.

Approach

To allow the trees comparison and implement the different use cases, we are proposing a group of interactive visualizations and have successfully implemented some of them. Our objective is to represent as much information of the tree differences as possible, including:

- Created and removed nodes.
- Absolute and relative differences of the node attributes values.
- Cardinality of the differences.

This characteristics can be enabled or disabled from the control panel, allowing the users to focus on the type of differences that interest them the most, and reduce or increase at their will the complexity of the visualization.

Limitations

To calculate the node attributes differences, we assume the hierarchies to contain an uniquely identifier for each node, that allow us to make a matching between the trees, and calculate the difference. We have established out of the scope of this project problems like finding matching subtrees or partial matchings.

In the same way of thinking we have chosen to work only on unordered trees, to simplify the problem domain.

Future Work

We have implemented a prototype of TreeVersity that implements the first use case. We believe that it could be an effective tool for the task at hand and plan to run experiments to test this theory.

Moreover we are working on identifying the best practices to represent and describe the type of differences that can be found in trees, as previously described. For this, we are planning of running an initial exploratory survey, and expect to enrich our tool with it.

Finally we are looking for possible application domains and specific case studies in which we can help experts compare tree structures and hierarchies, if you have a dataset and a problem that fits this description, please contact us, we are eager to work with you.

Acknowledgements

We thank the Center for Integrated Transportation Systems Management (a Tier 1 Transportation Center at the University of Maryland) for partial support of this research. We also thank the Center for Advanced Transportation Technology Laboratory (CATT LAB), Michael Pack, Michael VanDaniker and Tom Jacobs for their suggestions and feedback.

Related Work

1. http://ww.smartmoney.com/map-of-the-market/
2. Tamara Munzner, François Guimbretière, Serdar Tasiran, Li Zhang, and Yunhong Zhou. *TreeJuxtaposer: Scalable tree comparison using Focus+Context with guaranteed visibility*. In ACM SIGGRAPH 2003 Papers, pages 453-462, San Diego, California, 2003. ACM.
3. John Alexis Guerra Gómez, Krist Wongsuphasawat, Taowei David Wang, Michael L. Pack, Catherine Plaisant. *Analyzing Incident Management Event Sequences With Interactive Visualization*. In Proceedings of the Transportation Research Board 90th annual meeting, The National Academies, Washington, DC (2011) [Published Version]