Visualization of Decision Trees based on General Line Coordinates to Support Explainable Models

Boris Kovalerchuk1, Alex Worland, Sridevi Wagle

Department of Computer Science, Central Washington University, USA

*Abstract***—Visualization of Machine Learning (ML) models is an important part of the ML process to enhance the interpretability and prediction accuracy of the ML models. This paper proposes a new method SPC-DT to visualize the Decision Tree (DT) models as interpretable models. These methods use a version of General Line Coordinates (GLC) called Shifted Paired Coordinates (SPC). In SPC, each n-D point is visualized in a set of shifted pairs of 2-D Cartesian coordinates as a directed graph. The new method expands and complement the capabilities of existing methods, to visualize DT models. It shows: (1) relations between attributes, (2) individual cases relative to the DT structure, (3) data flow in the DT, (4) how tight is each split threshold in the DT nodes, and (5) density of cases in parts of the n-D space. This information is important for domain experts for evaluating and improving the DT models, including avoiding overgeneralization, and overfitting of models, along with their performance. The benefits of the methods are demonstrated in the case studies, using three real datasets.**

Keywords— Visual Analytics, Interpretability, General Line Coordinates, Machine learning, Decision trees

# Introduction

Evaluation and enhancing of Machine Learning (ML) models including their interpretability and prediction accuracy requires advanced tools. Visualization is an important approach for this [4]. The paper deals with visualization of trained Decision Trees (DTs) as interpretable models offering a new visualization method SPC-DT.

The motivation for developing of this method is the following. The existing methods for visualizing decision trees like shown in Fig. 1 show only a general structure of DT. It does not show how the cases of different classes are split by the DT. It requires additional information to be put next to each note how many cases of each class are going to this node and and/or to put a small distribution diagram as shown in Fig. 2.

The representation in Fig. 1 allows to trace the classification prediction from the root to the leaf for a new n-D point to be predicted. However, it is not sufficient to tell how good this

prediction is. Can it be trusted? The answer needs to be derived from additional knowledge like the purity of each node, DT accuracy on training and validation data and so on. Fig. 2 from [Elzen, Wijk, 2011] shows a part of that information

|  |
| --- |
|  |
| Fig. 1. Traditional visualization of Wisconsin Breast Cancer (WBC) data in the decision tree. Green edges and nodes indicate the benign class and red edges and nodes indicate the malignant class. |

It is well known that the **thresholds** at nodes of the DT are **not unique**. Shifting thresholds within some limits can keep the DT prediction on training and/or validation data the same. However, a new borderline case with values near the thresholds can be vulnerable due to this variability of the thresholds. The prediction can dramatically change class, making it **unstable**.The existing visualizations of decision tree are weak to support analysis of such situations to ensure that the prediction will be stable and meaningful.

|  |
| --- |
| Graphical user interface, application  Description automatically generated |
| Fig. 2. DT visualization as streams [Elzen, Wijk, 2011]. |

A related problem is **tight thresholds**, where cases of opposite classes are close to the threshold and overlap. It is difficult to set up right thresholds automatically. A slight change of the threshold can change classification for many training, validation and new cases making the prediction unreliable like we already discussed above about new borderline cases. This for the tight borders many training and validation cases are borderline cases in contrast with the wide borders between classes.

Interaction with visuals has a significant advantage for adjusting thresholds and especially with the abilities to use the domain knowledge provided by the end users. The existing visualizations of DTs are weak in this.

The next well-known problem with DTs is **overfitting** that requires pruning DTs by shortening some branches. It cuts out the number of the thresholds but makes a DT less accurate. It is difficult to balance the size of the tree and accuracy. The visualization can help in setting up the right balance by incorporating the domain knowledge of the end users.

Some known visualizations accompany decision trees with frequency distributions of attributes at each node is a shown in Fig. 2, but it is difficult to scale for larger DTs.

Next the existing visualizations do not support showing all training data with the decision tree and similarly do not show all validation data with the decision tree to be able to compare how they are represented in the DT in detail. For instance, Fig. 2 does not allow to see which cases are borderline cases.

These limits the abilities for the deep analysis and improvement of the decision tree.

The proposed approach allows to visualize separately or together all training, validation, and test data in the decision tree. It is important to note that this visualization is lossless [9]. Thus, it will allow to see how validation and test data are related to the training data and judge how stable and reliable is evaluation of quality of the decision tree. It also allows to see outliers. Alex, you can make pictures of DT in WBC training data (90%) and separately on the validation data (10%) and add then to section3?

For the decision tree model which is already constructed the problem is to convince the user that the DT prediction should be **trusted**. Unfortunately, the existing visualizations are very limited in supporting the user to get such confidence. To address this issue the proposed approach allows **tracing** any new case relative to the training and validation data in the decision tree. It also helps to justify or adjust thresholds of the DT for borderline cases.

While overfitting is a well-known problem for the decision trees the **overgeneralization** is less noticed but also is very important [10]. Consider an example where attribute x1 is in the range [0,10] and a branch of the DT states that if *x*1≤2.5 then case belongs to class 1. Also, no case of class 1 can be less than 1.5 and all training cases are within that range [1.5, 2.5]. The decision three overgeneralized cases of class 1 to the interval [0, 2.5]. The proposed visualization method allows to see this overgeneralization and adjust the range to its actual limits interactively.

# II. Shifted Paired Coordinates – Decision Tree Method

## *Concept and Algorithm*

The concept of General Line Coordinates (GLC) [3,9] is behind the method. Specifically, the method employs the Shifted Paired Coordinates (SPC. As the name indicates, SPC is set of shifted pairs of Cartesian coordinates. In SPC, a directed graph (digraph) visualizes an n-D point **x,** in 2-D, with nodes formed by consecutive pairs (*xi*,*xi*+1) of values of coordinates of **x,** connected by directed edges.

The method expands and complement capabilities available in the traditional visualization of DT models by visualizing: (1) *relations* between attributes, (2) *individual cases* relative to the DT, (3) *detailed* *data flow* in the DT, (4) *how* *tight* is each split threshold in the DT nodes, and (5) *density* of cases in parts of the n-D space. This information has a significant value for domain experts in evaluating and improving DT models. To the best of our knowledge none of the existing DT visualization methods support all these capabilities.

The idea of the Shifted Paired Coordinates – Decision Tree (**SPC-DT)** method is presented below first conceptually and then in examples. The major steps of the SPC-DT process to generate DT visualization from a given decision tree model are:

* Parsing DT model.
* Pairing attributes.
* Building a set of paired Cartesian coordinates.
* Drawing each pair of coordinates in the default location

shifted one over another.

* Mapping a part of the decision tree to a respective pair of coordinates as rectangles based on the threshold values. Color rectangles according to classes.
* Drawing selected n-D points as directed graphs in the SPC.
* Interactive modification of visualization. The options include: (1) change a set of visualized n-D points, (2) change the location of pair of coordinates, (3) change colors of the classes, (4) negate/flip coordinates, (4) swap vertical and horizontal coordinates, (5) condense points in the rectangle, (6) show overall DT structure (7) adjust thresholds, (8) compute accuracy after adjusting thresholds.

The concept of **shifted paired coordinates** (**SPC**) is explained below using an example. Consider a 6-D point **x** = (*x*1,*x*2,…,*x*6) = (1,2,3,2,5,1). In SPC visualization, the first pair (*x*1,*x*2) = (1,2) is visualized as a point in the first pair of coordinates (X1,X2). The second pair (*x*3,*x*4)=(3,2) is visualized in the second pair of coordinates (X3,X4) and the last pair (*x*5,*x*6)=(5,1) is visualized in the third pair of coordinates (X5,X6).

These pairs of coordinates are shifted, relative to each other, to avoid overlap. Then these three points are connected forming a directed graph (1,2) → (3, 2) → (5,1). See Fig.3. In this way, shifted pairs of coordinates visualize any n-D point, in 2-D, without loss of information [3,9].

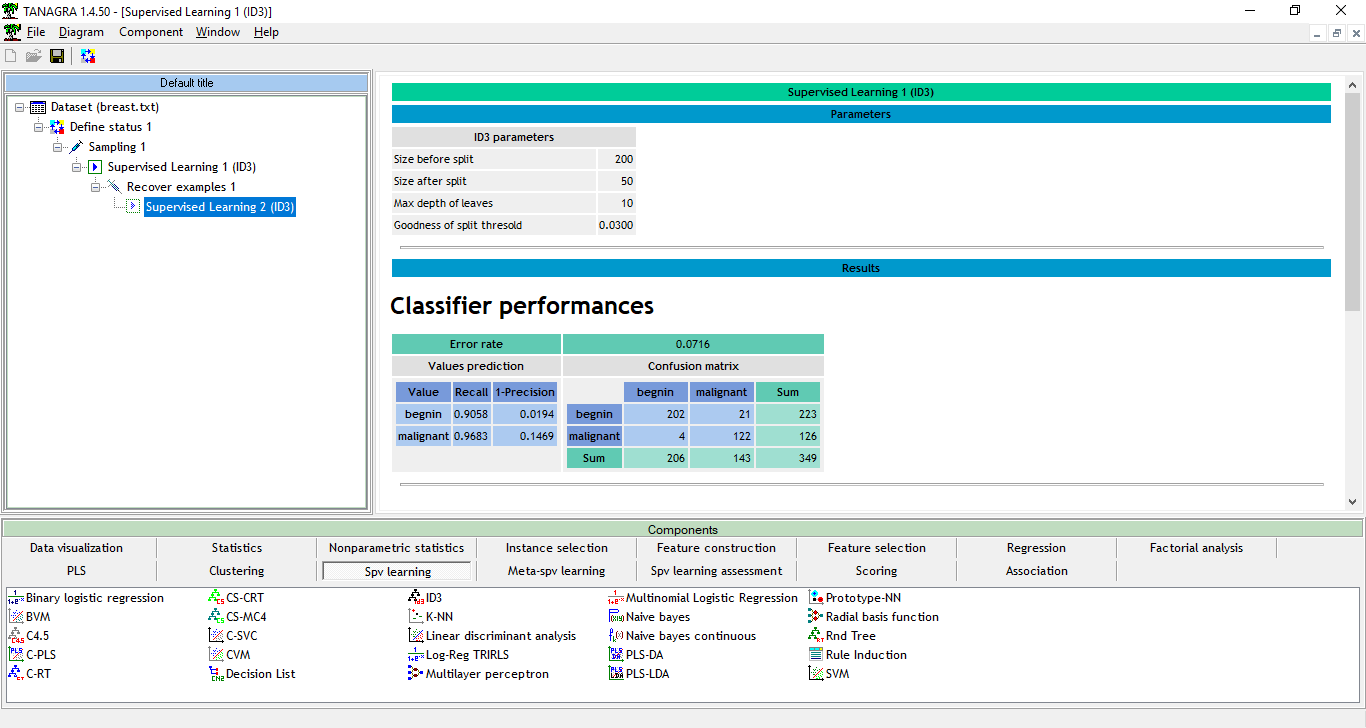
|  |
| --- |
| (3,2)  (5,1)  (1,2) |
| Fig. 3. 6-D point x = (x1,x2,…,x6) = (1,2,3,2,5,1) in Shifted Paired Coordinates (SPC) visualization. |

## *Example*

Below the SPC-DT method is explained, using the ID3 Decision Tree, trained on 349 cases of for Wisconsin Breast Cancer (WBC) data [Dua, Graff, 2019]. These cases are about 50% of all WBC Data. Fig. 4 presents this DT. Table 1 shows its performance and Fig. 1 shows the traditional visualization of this DT.

|  |
| --- |
| ucellsize < 2.5  bnuclei < 4.5 then class = **benign** (100.00 % of 200 cases)  bnuclei ≥ 4.5 then class = **malignant** (66.67 % of 6 cases)  ucellsize ≥ 2.5  bchromatin < 1.5 then class = **benign** (87.50 % of 8 cases)  bchromatin ≥ 1.5  clump < 4.5  bnuclei < 6.0  mgadhesion < 3.5 then class = **benign**  (100.00 % of 5 cases)  mgadhesion ≥ 3.5 then class = **malignant** (66.67 % of 6 cases)  bnuclei ≥ 6.0 then class = **malignant** (100.00 % of 8 cases)  clump ≥ 4.5 then class = **malignant** (93.97 % of 116 cases) |
| Figure 4. ID3 Decision Tree for WBC data. |

Table 1. ID3 DT performance.



While WBC data contain nine attributes, this decision tree uses only five attributes, where the attribute bnuclei is used twice with two different splits. Thus, we have 5-D data in this DT that we represent as a 6-D point with bnuclei repeated. This also helps to get three pairs of coordinates required for SPC.

The root of this tree is based on the attribute ucellsize, the next node is based on the attribute bnuclei. For SPC-DT visualization, these two attributes form the first pair of the Cartesian coordinates, with ucellsize as a horizontal coordinate, and bnuclei as a vertical coordinate. Figure 3 shows it on the left. The next nodes of the DT, based on bchromatin and clump attributes, form the second pair of coordinates, in the middle of Fig.5. The last nodes, based on bnuclei and mgadhesion attributes, form the last pair of the Cartesian coordinates, shown in Fig.5 on the right.

|  |
| --- |
|  |

|  |
| --- |
| Fig. 5. WBC decision tree in Shifted Paired Coordinates. Lighter green and red colors show lower density of cases in the respective areas. |

If the values of first two coordinates (ucellsize, bnuclei) of a 6-D point are in the grey area then it means that the decision tree did not define the class (label) of 5-D case **x** yet when the case is traced in the DT.

Similarly, a grey area, in the second pair of coordinates, is the area, where the decision tree did not define the class yet, too. The green areas, in each pair of coordinates, are areas where the DT has already determined that the class is benign. Similarly, the red areas are the areas, where the DT has already determined that the class is malignant. For example, the first two nodes of DT represent the rule:

If ucellsize < 2.5 & bnuclei < 4.5 then class = **benign**.

The green area in the first pair of coordinates visualizes this rule. Similarly, the red area in this pair of coordinates represents the second rule for this DT:

If ucellsize < 2.5 & bnuclei ≥ 4.5 then class = **malignant**.

The grey area in this pair of coordinates represents cases that the DT did not classify yet. The process of producing visualization is continued for the remaining nodes of the DT with respective DT rules:

If ucellsize ≥ 2.5 & bchromatin < 1.5 then class = **benign** (see the green arrow from the grey area in the first plot to the green area in the middle plot).

If bchromatin ≥ 1.5 & clump ≥ 4.5 then class = **malignant** (see the red arrow from the grey area in the first plot to the red area in the middle plot).

If bchromatin ≥ 1.5 & clump < 4.5 & bnuclei < 6.0 & mgadhesion < 3.5 then class = **benign** (see the green arrow from the grey area in the middle plot to the green area in the right plot).

If bchromatin ≥ 1.5 & clump < 4.5 & bnuclei < 6.0 & mgadhesion ≥ 3.5 then class = **malignant** (see the red arrow from the grey area in the middle plot to the red area in the right plot).

If bchromatin ≥ 1.5 & clump < 4.5 & bnuclei ≥ 6.0 then class=**malignant** (see the red arrow from the grey area in the middle plot to the upper red area in the right plot). \

Fig. 6 shows several 6-D points in DT that reach terminal nodes of the DT.

|  |
| --- |
|  |
| Fig. 6 WBC decision tree in SPC with several cases that reach terminal nodes of the DT. |

Advantages of SCP-DT visualizations in Figs 5. and 6 are as follows:

1. show relations between pairs of attributes inside of each plot,
2. show relations between 4 attributes using green and red arrows between two plots,
3. show relations between all 6 attributes using green and red arrows between three plots and more attributes for larger dimensions,
4. show the tree structure and data flow by arrows in the DT.

Arrows can represent individual cases like in Fig.6. Graphs of individual cases in Fig. 6 allows to see how close these cases to each split threshold in the DT nodes. Graphs for training cases allows to see the density of the cases in the parts of the n-D space, Using color intensity (lighter/darker) shows the density of cases in the parts of the n-D space (see Fig. 5).

All these features together help to get confidence in the DT model interpretability and accuracy for the future predictions. For instance, if all actual training cases are far from the threshold (border) area, but a new case to be predicted is close to this threshold, then DT model is likely overgeneralized training data making the predictions too risky.

In general, SPC-DT visualization supports domain experts for evaluating the DT model performance, interpretation, overgeneralization and overfitting.

The interactive version of SPC-DT method supports turning on/off “context” attributes (attributes that are not part of the tree), but a part of the given dataset. These attributes are associated with the tree root. When such attributes are turned off, they are greyed or made semitransparent that still allows seeing the context.

To minimize occlusion and simplify visual patterns when visualizing a large number of cases and classes SPC-DT method allows showing only graphs of “centers”, min and max of clusters of cases with the remaining cases grey or semitransparent. Next, to help the user to make prediction and be confident in it SPC-DT method allows showing the error rate for each dataset, branch of the three and the full confusion matrix.

# III Case studies <work here>

Below visualization experiments with SPC-DT on two other datasets from UCI Machine Learning Repository [Dua, Graff, 2019] are reported.

## *A. Dataset 1: Wisconsin Breast Cancer (Original)*

This dataset contains 748 cases with 4 attributes each. Figure 8 shows visualization of the DT presented below using SPC-DT method.

Decision tree

|  |
| --- |
| Diagram  Description automatically generated |
| Fig. 7. WBC Decision Tree |

Fig. 7 shows all WBC data in SPC-DT in default locations for each coordinate pair. The gray areas indicate locations within each paired coordinate space where a case’s class cannot be determined at the current DT level and must be followed to the next DT level to be assessed further. The red areas indicate cases which are determined by the DT to be of class “malignant” while the green areas indicated cases which are determined by the DT to be of class “benign.” These cases terminate at the current level since a class has already been determined.

This visualization is not very informative due to the heavy overlap of lines. This makes following individual cases difficult, especially in attribute pairs with multiple gray zones. Figs. 8 and 9 show the same data with the individual coordinate pairs are rearranged and relocated in order to make the SPC-DT visualization more easily interpretable.

|  |
| --- |
|  |
| Fig. 8. WBC in SPC-DT with Rearrangement and Relocation |

Fig.8 shows the same data as Fig. 7, but with a significant improvement in interpretability due to the rearrangement and relocation of coordinate pair plots. Care has been taken to reduce the amount of line overlap, which greatly increases the interpretability of the DT and the WBC cases that pass through it.

However, more can be done to make the SPC-DT visualization more interpretable. Many coordinate pair plots still suffer from line overlap and an additional problem can be addressed by minimizing the number of cases that pass over classes of which that case does not belong.

|  |
| --- |
|  |
| Fig. 9. WBC in SPC-DT with Rearrangement, Relocation, and Condensation |

Fig 9. shows the same data as Figs. 7 and 8, but with the condensation of points in gray zones. This allows for a significant increase of interpretability, especially in coordinate pair plots that have a high amount of cases passing through them like plotId: 1 (x: clump, y: ucellsize), as well as plots that have a high number of gray zones like plotId: 3 (x: bnuclei, y: clump).

However, the use of condensation naturally reduces the amount of data that can be losslessly visualized. The closer a point gets to the border of classification zones, the more uncertain the classification can be and by condensing all points in a 2D space to a single 1D point, this uncertainty information is lost.

## *B. Dataset 2: Iris Dataset*

This dataset contains 150 cases of 3 classes. Each case is presented by 4 attributes. Figure 10 shows visualization of the DT described below using SPC-DT method.

Decision tree

## *C. Dataset 3*

# IV. Comparison and Conclusion

About 200 different DT visualization were reported in 2011 [Schulz, 2011], and now Treevis.net shows over 300 of them. While these are impressive numbers, only a few of these DT visualizations are detailed enough to represent the machine learning models. For instance, a Treemap DT visualization, which is very powerful for showing large complex trees, does not show split thresholds, in the DT machine learning models.

Typically, alternative DT visualizations in machine learning show the number of cases in each node and/or one-dimensional (marginal) distribution of the cases.

One of the major advantages, of the SPC-DT method, is in its ability to shows the actual cases, and their 2-D distributions, in the (Xi,Xj) coordinates, presented in the adjacent nodes of the DT. In particular, it allows tracing of the misclassified cases to improve the accuracy of the DT model interactively.

For instance, when misclassified cases are close to the threshold of some nodes, then a user can interactively move the thresholds in these nodes in the SPC-DT visualization to change classification of those cases, and see how it affects the classification of other training and validation cases.

In SPC-DT method, a user has abilities interactively optimize the DT. While some other DT visualization such as [Parr, Glower, 2019] allow it by using 1-D distributions, SPC-DT allows it in more detail in 2-D distributions.

In addition, SPC-DT provides a more compact visualization of cases with the less number of tracing points. For instance, in Fig. 5 and 6, tracing the 6-D case (5, 8, 3, 4, 6) would require 3 tracing points. In contrast, in Figs. 1 and 2 it would require 5 tracing points.

This paper proposed a new DT visualization method for machine learning. The features and advantages of these methods include showing relations between multiple pairs of attributes, in each plot, individual cases, closeness of the cases to the split thresholds, in the DT nodes, and the density of the cases, in the parts of the n-D space. These features together help in getting confidence in the DT model interpretability and accuracy for the future predictions, avoiding overgeneralization and overfittings.

# References

1. Dua, D. and Graff, C. UCI Machine Learning Repository, http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science, 2019.
2. Elzen, van den, S. J., & Wijk, van, J. J. BaobabView: Interactive construction and analysis of decision trees. In Proceedings IEEE Symposium on Visual Analytics Science and Technology, VAST 2011, pp. 151-160, IEEE, 2011, DOI: 10.1109/VAST.2011.6102453
3. Kovalerchuk, B., Visual Knowledge Discovery and Machine Learning, Springer Nature, 2018.
4. Ming Y, Qu H, Bertini E. RuleMatrix: visualizing and understanding classifiers with rules. IEEE transactions on visualization and computer graphics. 2018 Aug 20;25(1):342-52.
5. Parr T., Grover P., How to visualize decision trees, https://explained.ai/decision-tree-viz/, 2019
6. R2D3, A visual introduction to machine learning, http://www.r2d3.us/visual-intro-to-machine-learning-part-1/, 2019
7. SAS: Working with Decision Trees, http://support.sas.com/documentation/cdl/en/vaug/68027/HTML/default/viewer.htm#n0q3i0zwng79kin1kb1zvpo9k312.htm, 2016.
8. Schulz, H. Treevis.net: A tree visualization reference. IEEE Computer Graphics and Applications, 31(6):11–15, Nov 2011. doi: 10.1109/MCG.2011
9. Kovalerchuk B, Grishin V. Adjustable general line coordinates for visual knowledge discovery in n-D data. Information Visualization. 2019 18(1):3-2
10. Kovalerchuk B, Grishin V. Reversible data visualization to support machine learning. In: International Conference on Human Interface and the Management of Information 2018, 45-59. Springer