

Orange and Visual Programming

Orange Widgets

Orange is a comprehensive, component-based framework for machine learning and data mining. For explorative data analysis, it provides a visual programming framework with emphasis on interactions and creative combinations of visual components — widgets.

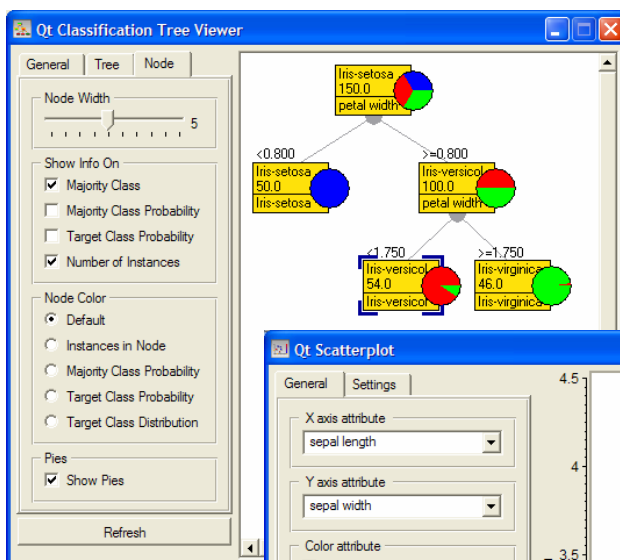
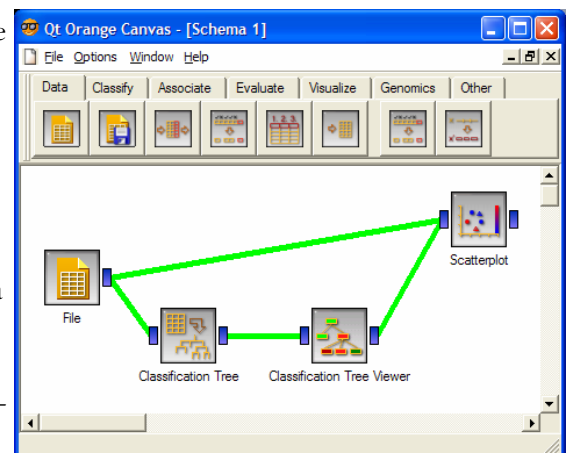
The power of widgets is in their modularity. Widgets can be connected through channels and communicate with each other by sending and receiving data tokens. Communication channels are typed and the system establishes the proper data connections automatically.

A particular set of connected widgets is called a schema. Orange schemas can be either set in Python scripts, or, preferably, designed through visual programming in an application called Orange Canvas.

Visual Programming in Orange Canvas

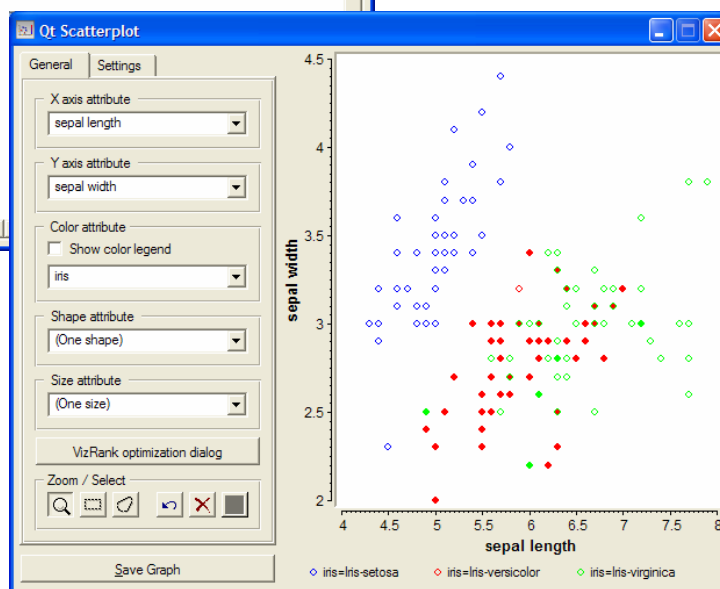
The programming process – creating a schema with widgets and their connections – is done visually through an easy-to-use graphic interface. A figure on this page shows Orange Canvas with a simple

schema. We have loaded the famous Iris data set (File widget), plotted the data in the scatterplot and build a classification tree.



But notice a small and useful trick. Classification tree visualization allows user to

click on the node, sending out the data that corresponds to the node. We have connected the output of tree visualization widget to the “Example Subset” channel of the scatterplot. This now displays the selected examples with filled circles to distinguish them from other examples in the data set. Selecting the nodes in decision tree updates the scatterplot immediately, turning out particular schema into nice and simple data exploration



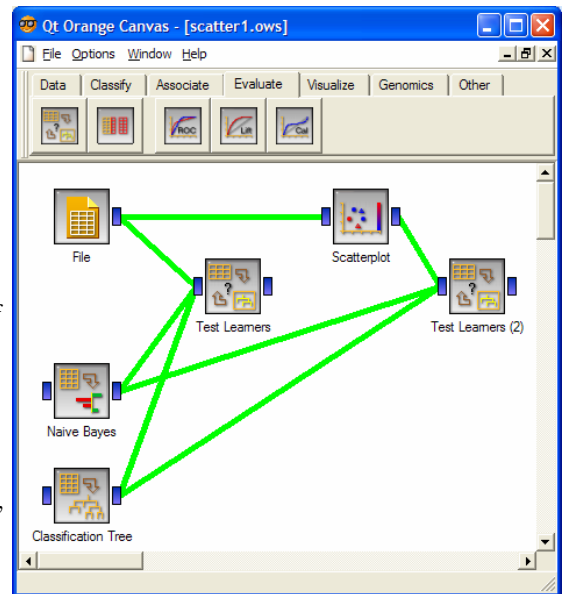
One More Schema: Making Classifiers Sweat

Interaction is a powerful concept behind Orange widgets. Not just connecting them to visually create application of your needs, widgets were carefully designed to present data and support interaction with visualization (selecting a data subset, an interesting part of the prediction model, ...).

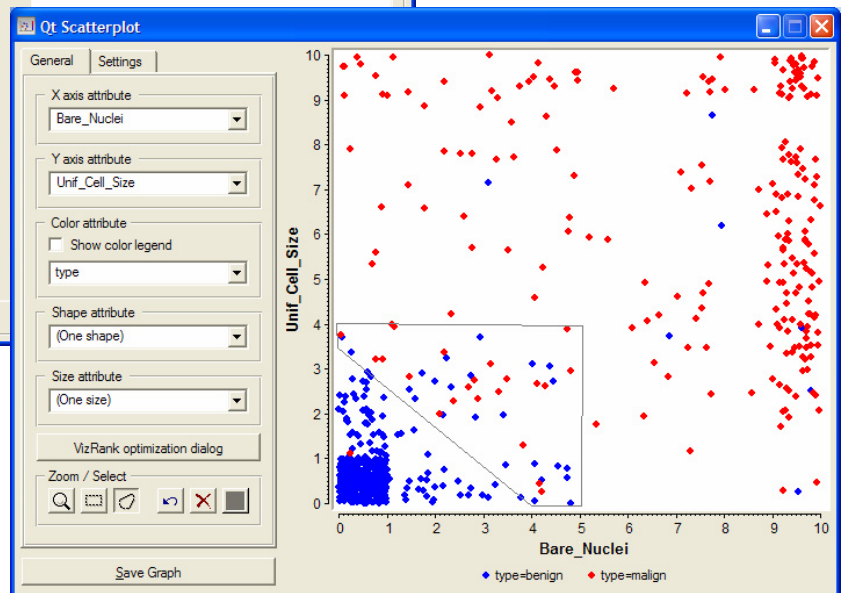
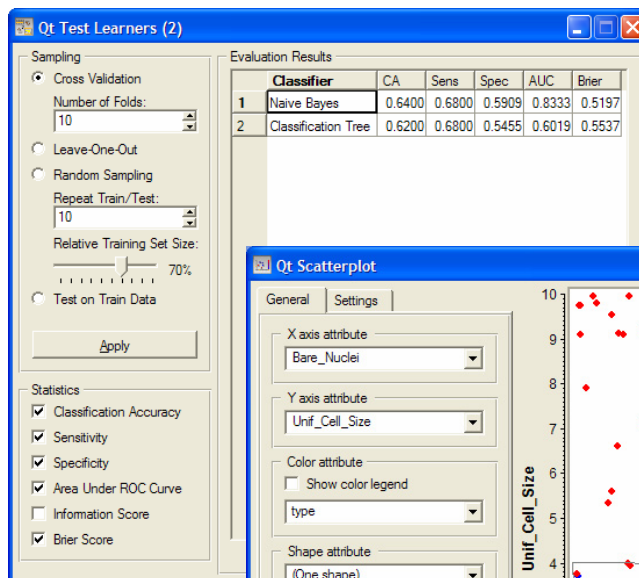
For illustration, here is one more example that rides on the power of interaction. The widget schema shown loads the data (Wisconsin Breast Cancer data set from the UCI ML repository) and uses scatterplot's wizard to show the visualization that best discriminates between cases. We also feed the data to the widget that using 10-fold cross validation compares the performance of naive Bayes and classifi-

cation tree algorithm. While they perform great, we are interested if there is any difference if we use the same algorithms on a harder data set. For this, we have selected a subset of data (scatterplot), fed it into the same evaluation schema, and — as expected — observed much poorer performance.

Like in the previous example, any change in selection of a data subsets issues the new token that gets propagated through the schema.



Widgets update their visualizations automatically, so one can easily explore how does the particular data selection affect the performance of the classifiers.



Orange widgets and Canvas are in beta release (Sep 2004). About fifty widgets have been implemented, support various tasks from data preprocessing, visualization and modelling. Check Orange's web page www.ailab.si/orange for updates.