BIONUMERICS Tutorial:

Building automated decision and action workflows

1 Introduction

Decision networks in BIONUMERICS should be seen as a construction kit that allows you to build your own automated decision or action work flows, with practically endless possibilities. They can be used to make decisions, predict features, perform queries, fill in fields, create graphs and plots, and much more. In this tutorial you will learn how to create these work flows in BIONUMERICS.

2 Preparing the database

The **DemoBase Connected** will be used in this tutorial and can be downloaded directly from the *BIONUMERICS Startup* window or restored from the back-up file available on our website:

1. To download the database directly from the *BIONUMERICS Startup* window, click the ⬇ button, located in the toolbar in the *BIONUMERICS Startup* window. Select **DemoBase Connected** from the list and select *Database* > *Download* ( ). Confirm the download action.

2. To restore the database from the back-up file, first download the file DemoBase_Connected.bnbk from https://www.applied-maths.com/download/sample-data, under 'DemoBase Connected'. In the *BIONUMERICS Startup* window, press the button, select *Restore database*, browse for the downloaded file and select *Create copy*. Specify a name and click <*OK*>.

✎ In contrast to other browsers, some versions of Internet Explorer rename the DemoBase_Connected.bnbk database backup file into DemoBase_Connected.zip. If this happens, you should manually remove the .zip file extension and replace with .bnbk. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the .zip file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

3 Creating a new decision network

As an example, we will create adecision network that discriminates between the three genera in the demobase **DemoBase Connected**, based upon the 16S rDNA sequences.

1. In the *BIONUMERICS Startup* window, double-click on the **DemoBase Connected** database to open it.

2. In the *Database entries* panel of the *Main* window, select all entries except those defined as STANDARD: select **Ctrl+A** to select all entries and unselect the entries defined as STANDARD by unchecking the check boxes next to the entries.

In the default configuration of the *Main* window, the *Decision networks* panel is seen as a tab behind the *Identification projects* panel.

3. Click on the header of the *Decision networks* panel to bring the panel to the top (Figure 1).
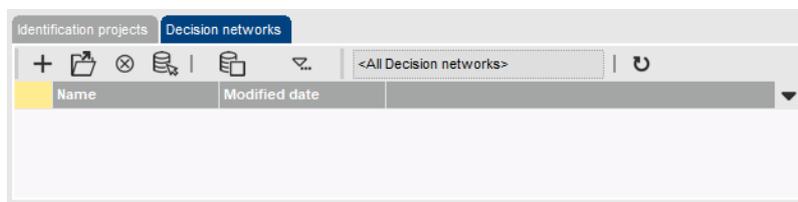
Figure 1: The *Decision networks* panel in the *Main* window.

4. To create a new empty decision network, highlight the *Decision networks* panel and select **Edit > Create new object...** (+).

5. In the *Create new object* dialog box that pops up, enter a name for the decision network (e.g. **MyDN**) and press <**OK**>.

The new decision network is now listed in the *Decision networks* panel and the decision network is opened, containing the current selection of entries.

4 Building a decision network

1. In the newly created decision network, open the **Data sources** group in the *Operators* panel by clicking on its ⊞ icon.

2. Double-click on **Sequence**, which opens a *Decision network operator* dialog box.

3. Press <**OK**> to edit the node properties for the sequence input node.

4. Enter e.g. "16S" as a **Name**, select **16S rDNA** (the only sequence type available in **DemoBase Connected**) as **Sequence type** and press <**OK**> (see Figure 2).

The network now contains one node, i.e. "16S".

5. If you click on an entry in the *Entry data* panel, the node and the *Node properties* panel are updated with the sequence data of the highlighted entry.

6. Select the node "16S" in the network (a selected node is bordered by a red line).

7. Open the **Sequence operators** group in the *Operators* panel and double-click on **Find subsequence**.

Figure 2: Sequence operator.

8. Press <***OK***> to edit the node properties.

9. Enter as ***Name*** "Signature 1", and as ***Comment*** "Recognizes Ambiorix", and check ***Use as output***.

10. Enter `GGGTGTAG` as ***Match sequence***, with zero mismatches allowed (see Figure 3).

Figure 3: Find subsequence operator.

11. Press <***OK***> to confirm the node properties.

The network is now ready to produce a first result.

12. In the *Decision Network* window, press the ▷ button to calculate the network.

The *Entry data* panel now contains one output column, ***Signature 1***, showing a boolean TRUE or FALSE for each entry. All *Ambiorix* entries have the boolean TRUE, the others FALSE.

In the *Network* panel (Figure 4), the node "Signature 1" is marked with a green flag, indicating that it is an output node, resulting in a column in the *Entry data* panel. The percentage of true and false entries is indicated as a green and red bar, respectively. In addition, the percentage of selected entries is indicated with a blue bar.
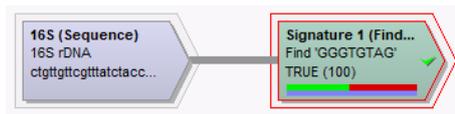
Figure 4: Simple decision network with one boolean output.

13. Select the node "16S" in the *Network* panel.

14. Open the **Sequence operators** group in the *Operators* panel and double-click on **Find subsequence**. Press <**OK**>.

15. Enter "Signature 2" as **Name**, "Recognizes Vercingetorix" as **Comment**, and `CGATCTCACG` as **Match sequence**, with zero mismatches allowed.

16. Check **Use as output** and press <**OK**> (see Figure 5).
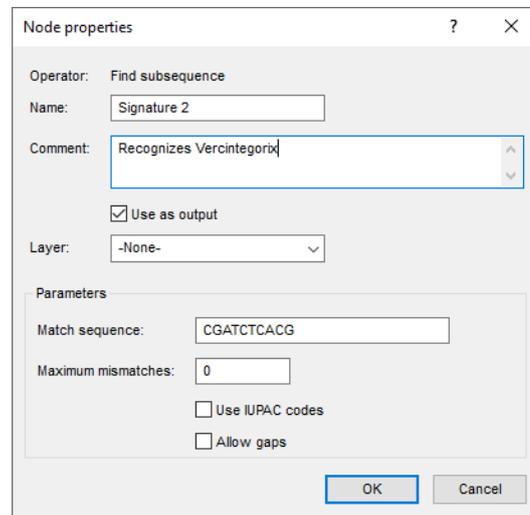
Figure 5: Find subsequence operator.

17. Press the ▷ button to calculate the network. The network now looks as in Figure 6.
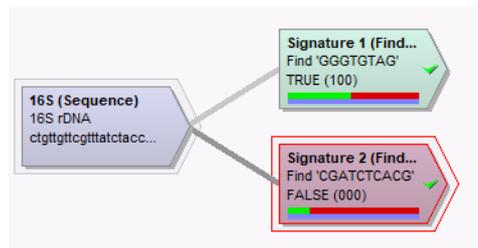
Figure 6: Decision network with two boolean output nodes.

A second column, **Signature 2**, is added to the *Entry data* panel. The *Ambiorix* entries have "Signature 1" TRUE and "Signature 2" FALSE, whereas the *Vercingetorix* entries have "Signature 2" TRUE and "Signature 1" FALSE. Entries of *Perdrix* have both signatures FALSE.

Although this type of network might allow you to predict the properties for new and unknown entries, the output is not very descriptive or easily interpretable. We will now turn the network into a more descriptive result.

This time we will make use of a Duplicator operator. This operator duplicates a node in the network, i.e. one output parameter from it, which can be chosen. This tool is useful if a node is to be used in more than one independent operations in the network. Theoretically one could branch the different operations from the same node, but the network could easily become unsurveyable. As we need to check "Signature 1" and "Signature 2" to be both negative for *Perdrix*, we will duplicate both booleans and branch the new operation from there.

18. Select the subsequence search node "Signature 1" and, in the *Operators* panel, double-click on the **Duplicator** operator in the **Miscellaneous** group.

19. Select "Signature 1 (IsMatch)" as the parameter to be duplicated and press <**OK**>.

20. Leave the **Name** and **Comment** fields blank and leave **Use as output** unchecked, as this is an intermediate node. Make sure **Hide link** is checked and press <**OK**>. This will place the node on a new line, separated from the other operators.

21. Select the duplicator node (red border) in the *Network* panel and double-click on the **NOT** operator in the **Boolean operators** group in the *Operators* panel. Press <**OK**>.

22. Enter "No Signature 1" as **Name** for this node and press <**OK**>.

23. Select the subsequence search node "Signature 2" and, in the *Operators* panel, double-click on the **Duplicator** operator in the **Miscellaneous** group.

24. Select "Signature 2 (IsMatch)" as the parameter to be duplicated and press <**OK**> twice.

25. Select the new duplicator node (red border) in the *Network* panel and double-click on the **NOT** operator in the **Boolean operators** group in the *Operators* panel. Press <**OK**>.

26. Enter "No Signature 2" as **Name** for this node and press <**OK**>.

27. Select both boolean nodes "NOT" by clicking the first and then, while holding down the **Ctrl**-key, clicking the second. Both nodes are now bordered in red.

28. In the *Operators* panel, double-click on the **AND** operator in the **Boolean operators** group. Press <**OK**>.

29. Enter "Perdrix" as **Name** for this node and press <**OK**>.

The network now looks as in Figure 7.

We now already have one boolean node called "Perdrix"; we still need to create similar nodes for the two other groups.

30. Select both nodes "Signature 1" and "No Signature 2" and combine them with a boolean operator AND.

31. As **Name** for this node, enter "Ambiorix" and press <**OK**>.

32. Select both nodes "Signature 2" and "No Signature 1" and combine them with a boolean operator AND.

33. As **Name** for this node, enter "Vercingetorix" and press <**OK**>.

The network now looks as in Figure 8. This network contains cross-branching operators, but is still surveyable thanks to the duplicator nodes.

34. In the *Decision Network* window, press the ▷ button to calculate the network.

The network now contains 3 mutually exclusive boolean nodes (only one out of the three can be true). Such nodes can be combined into a categorical set, i.e. a set of multiple states (categories),
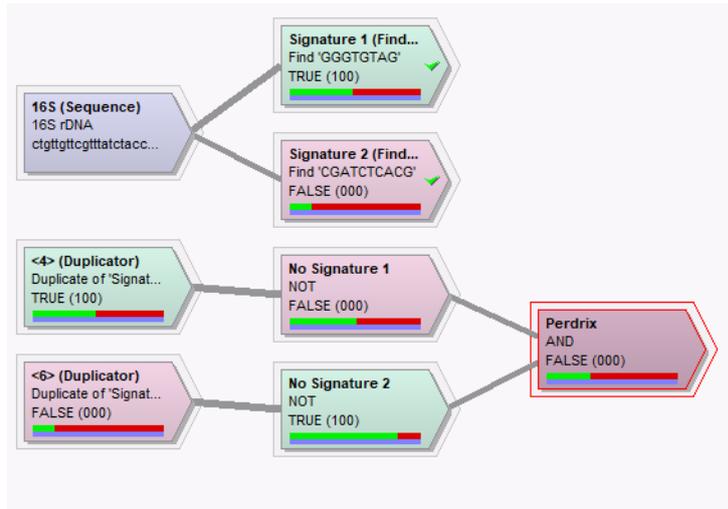
Figure 7: Decision network with duplicated nodes.
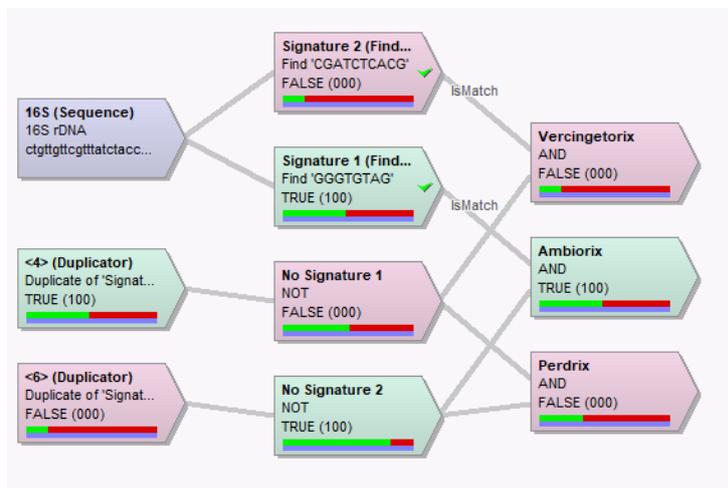
Figure 8: Decision network leading to three categorical boolean nodes.

of which one state is true for each individual. We will combine these categories into one node that tells the name of the genus:

35. Using the **Ctrl-** key, select all three end nodes ("Vercingetorix", "Ambiorix" and "Perdrix") and create a new node using the **Categorical combiner** from **Boolean operators** group in the *Operators* panel. Press <**OK**>.

36. Enter "Genus" as **Name** and check **Use as output** (see Figure 9).

37. Press <**OK**> to confirm the node properties.

38. In the *Decision Network* window, press the ▷ button to calculate the network.

The *Entry data* panel now contains a new output column, **Genus**, showing the genus name for all entries selected in the decision network.

Finally, we will add an output action operator to write the result of the network in a database field.

39. Save the changes in the *Decision Network* window with **File** > **Save** (🖫) and close the window with **File** > **Exit**.
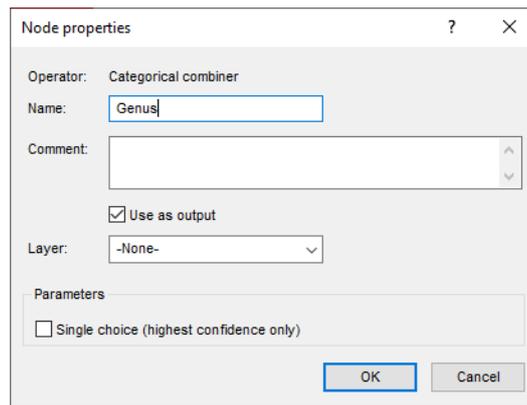
Figure 9: Categorical combiner operator.

40. In the *Main* window, add a new information field: select **Edit** > **Information fields** > **Add information field...**, specify a name (e.g. **Name by DN**) and press <**OK**> to save the field to the database.

41. Make sure all entries are selected except those defined as STANDARD, and double-click on the decision network displayed in the *Decision networks* panel.

42. In the *Decision Network* window, select the categorical combiner node "Genus" and double-click the **Write to field** operator from the **Output actions** group in the *Operators* panel. Press <**OK**>.

43. Enter a name, e.g. **Write to database** and select the field **Name by DN** as the **Database field name**. Press <**OK**>.

The finished decision network now looks as in Figure 10. The node that perform an output action is orange, to indicate that this node can perform changes to the database. For safety reasons, the output actions are not executed automatically when the network is calculated using the ▷ button.
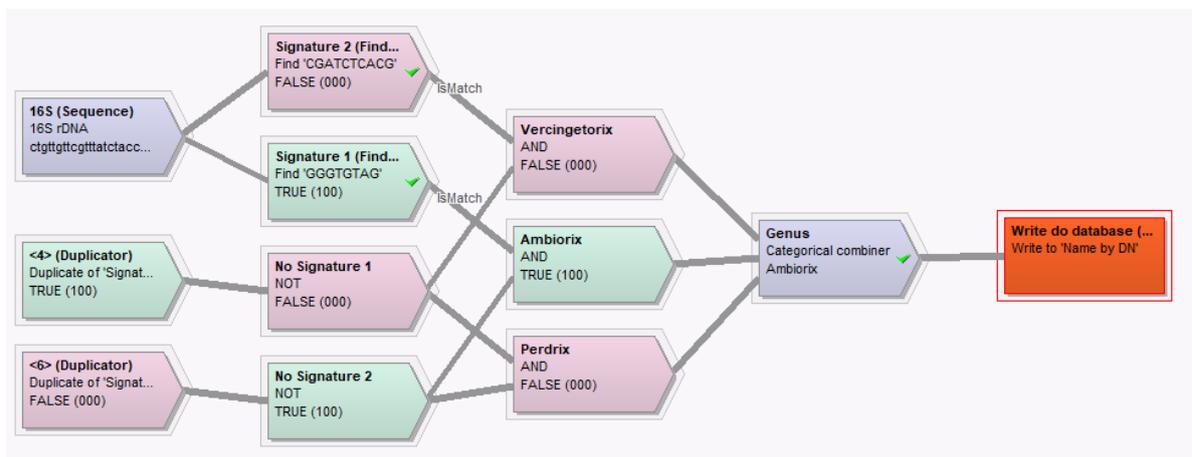
Figure 10: Decision network that decides between three groups and writes the output to the database.

44. To calculate the network and execute the output action(s), press the !▷ button.

To alert you that the network will now perform changes to the database, a warning box appears (Figure 11).

45. Press <**OK**> to confirm the execution.
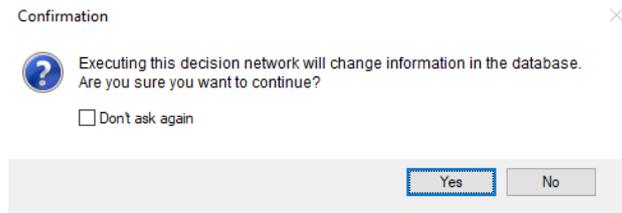
Figure 11: Warning box that appears if a network containing output nodes is executed.

When finished, the database field ***Name by DN*** contains the genus names as defined by the decision network.

5 Creating charts from a decision network

BIONUMERICS can plot the results of a decision network in a chart (scatter plot, bar graph, contingency table, etc.) by using its chart and statistics tools.

1. Recalculate the network with ▷ .

2. Select the "Genus" node and select ***Network*** > ***Plot in chart window...***.

A bar graph appears, showing the relative occurrences of the three genera (Figure 12).
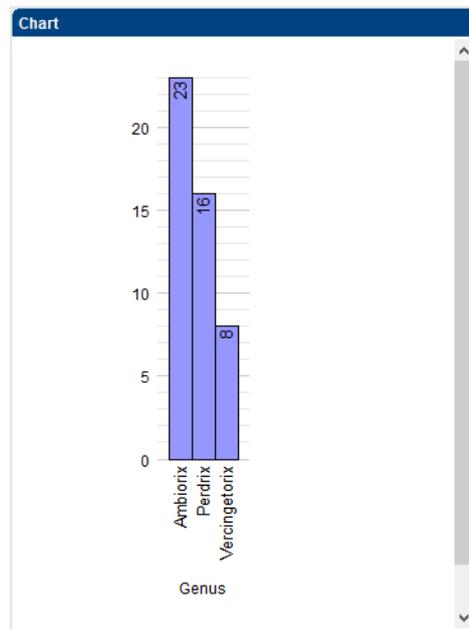
Figure 12: Bar graph popped up from categorical combiner output node.

3. Close the *Chart & statistics* window with ***File*** > ***Exit***.

To create a graph each time the network is executed proceed as follows:

4. Select the "Genus" node, and create a new output action node by double-clicking on the ***Histogram*** operator in the ***Charts*** group under ***Output actions*** in the *Operators* panel. Press <***OK***> twice.

5. Execute the network by pressing the ⏵ button.

A bar graph is created and the information in the database field **Name by DN** is updated.

6. Close the *Chart & statistics* window with **File** > **Exit**.

7. Save and close the *Decision Network* window.