

# Appendix A

## The CohesionOptimizer Tool

CohesionOptimizer is a GUI-enabled software tool that allows the user to automatically obtain clustering tree visualizations of Java functions in the form of dendrograms. The tool has been entirely coded in Java. This section describes the various features of the tool. Subsection A.1 elaborates on how to use the tool. Subsection A.2 gives the minimum system requirements of the tool.

### A.1 Using the Tool

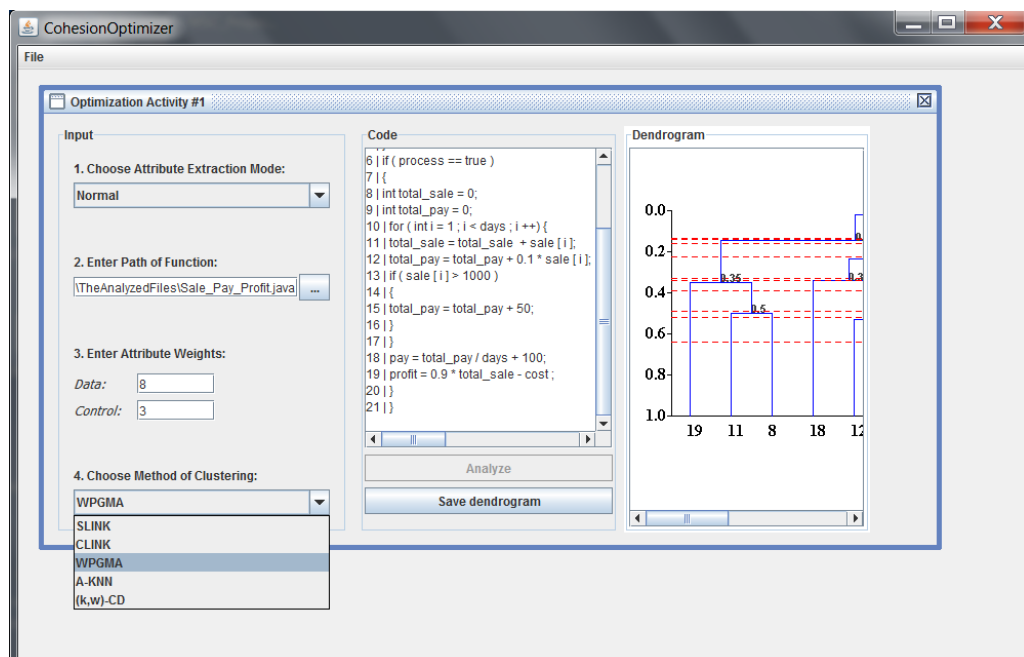


Figure A.1: The CohesionOptimizer software tool.

In order to execute the application the user has to open the .jar application file of the software. Then in order to perform a clustering of a function, the user will have to open a new “Optimization Activity” from the File menu. The window shown in Fig. A.1 would then appear.

### A.1.1 Input

The user will have to input the following details in the “Optimization Activity” window,

- the attribute extraction mode,
- the weight ratio between the data and control attributes,
- the pathname of a .java file which contains the function to be restructured,
- the method of clustering that is to be used.

**Format of Input File.** In order to interpret<sup>1</sup> the input function, the tool expects the code of the function to follow traditional formats: Each statement of the code must exist in a single line. The first line of the code must be the name of the function with an opening brace and must be tagged with the text, “#HEADER#”. Declaration statements must be tagged “#DEC#”. Each control statement must start from a new-line. (E.g., the code “...} else {” must be separated into two lines, i.e., “...}” in one line and “else {” in the other.). Single-line comments, blank lines may be present.

**Attribute Extraction Mode.** Two attribute extraction modes are available: The Normal Attribute Extraction Mode directly follows the attribute selection criteria of Lung *et al.* [LXZS06] and Selective Attribute Extraction Mode, which follows the new attribute selection strategy presented in this thesis. For obtaining best results while using  $(k, w)$ -CC, we recommend the Selective Attribute Extraction Mode.

**Weight Ratio.** For best results, weight ratios of 8:3, 5:2, 3:1 are recommended for data to control attributes respectively.

**Clustering method.** Five clustering algorithms can be used to perform clustering: Single Linkage Algorithm (SLINK), Complete Linkage Algorithm

---

<sup>1</sup>A custom-built parser based on the Java StringTokenizer class was used for this purpose.

(CLINK), Weighted Pair-Group Method of Arithmetic Averages (WPGMA), Adaptive k-Nearest Neighbour Algorithm (A-KNN), and  $(k, w)$ -Core Clustering ( $(k, w)$ -CC).

After inputting the above information, the “Code” box shall display an enumerated version of the function, where each statement/LOC of the function is prefixed with a unique number. Unique numbers are given only to non-comment lines of code.

### A.1.2 Output

The user can then generate a dendrogram for the input java function by clicking on the “Analyze” button. The dendrogram is displayed inside the “Dendrogram” box. The dendrogram displays the entities of the function in the horizontal axis, on a vertical scale of similarity ranging from 0 to 1. The blue lines indicate the clusters. The red-dashed lines indicate the possible lines of cut.

The “Save Dendrogram” button saves the output dendrogram as a .png image file in the same directory where the software tool application file (i.e., the .jar file) is located.

## A.2 System Requirements

CohesionOptimizer is a cross-platform application that can be used in both Linux and Windows based machines with Java enabled. For best user experience we recommend the application to be used in a system configured with at least the following settings,

- Operation System: Windows XP (for Windows-based systems), Ubuntu 9.04 (for Linux-based systems)
- Processor: 1.4 GHz
- RAM: 1 Gb