Information Modelling & Analysis – Project 1

Claudio Maggioni

Code Repository

The code and result files part of this submission can be found at:

Repository: https://github.com/infoMA2023/project-01-god-classes-maggicl

Commit ID: **TBD**

Data Pre-Processing

God Classes

Table 1: Identified God Classes

Class Name	# Methods
org.apache.xerces.dom.CoreDocumentImpl	125
org.apache.xerces.impl.xs.traversers.XSDHandler	118
org.apache.xerces.xinclude.XIncludeHandler	116
org.apache.xerces.impl.dtd.DTDGrammar	101

The god classes I identified, and their corresponding number of methods can be found in Table 1.

Note: the number of methods was computed by considering only methods and not constructors.

Feature Vectors

Table 2 shows aggregate numbers regarding the extracted feature vectors for the god classes.

Table 2: Feature vector summary (*= used at least once)

Class Name	# Feature Vectors	# Attributes*
...	...	...

Clustering

Algorithm Configurations

Report/comment the algorithm configurations (distance function, linkage rule, etc.). You may do so in any form you feel suited, but a short paragraph of text is probably sufficient.

Testing Various K & Silhouette Scores

(1) Report data about the clusters produced by the two algorithms at various k (#clusters, size of clusters, silhouette scores). You may use any suitable format (table, graph, ...).

(2) Briefly comment your results. What is the best configuration, and why? Anything else you observed?

Evaluation

Ground Truth

I computed the ground truth using the command .... The generated files are checked into the repository with the names ....

Comment briefly on the strengths & weaknesses of our ground truth.

Precision and Recall

Table 3: Evaluation Summary

Class Name	Agglomerative		K-Means	
	Prec.	Recall	Prec.	Recall
...	...	...	...	...

Precision and Recall, for the optimal configurations found in Section 3, are reported in Table 3.

Practical Usefulness

Discuss the practical usefulness of the obtained code refactoring assistant in a realistic setting (1 paragraph).