

An Active Learning Method for Software Effort Estimation

Dataset	Percentage
Coc81	52.38%
Nasa93	69.84%
Desharnais	40.74%
Albrecht	54.17%
Finnish	50.00%
Kemerer	33.33%
Maxwell	46.77%
Miyazaki	43.75%
Telecom	38.89%

Figure 1: Percentage of instances that are not closest neighbor to any other instance.

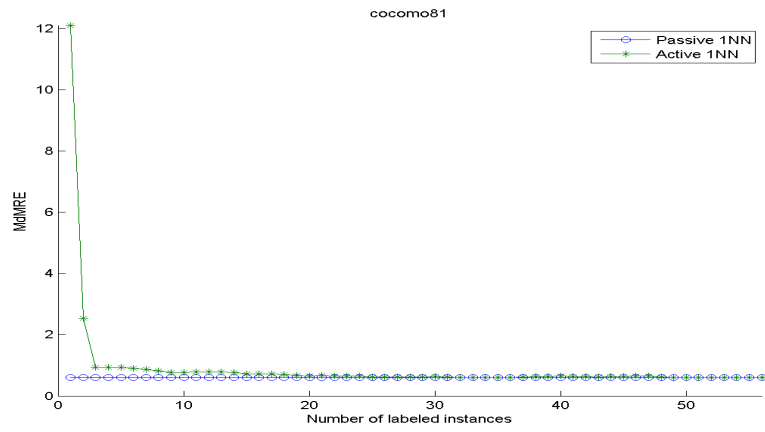
ABSTRACT

Proposed is an active learning strategy based on the *popularity* of instances in the datasets.

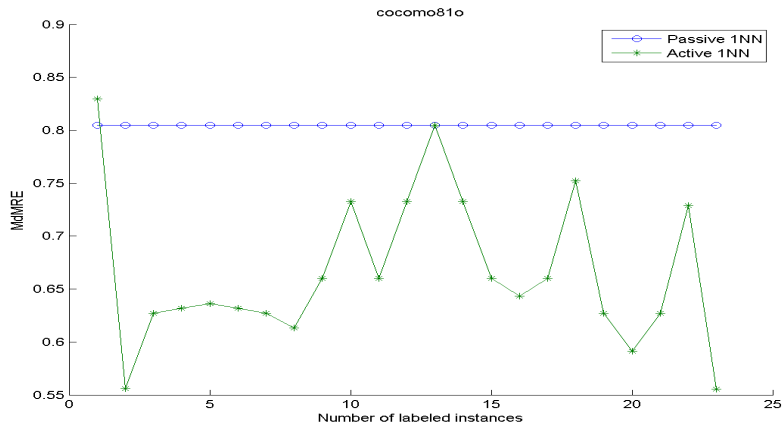
Assumption is that the hypothetical-expert is able to provide perfect labels, i.e. he can perfectly estimate the effort of the queried project.

Categories and Subject Descriptors

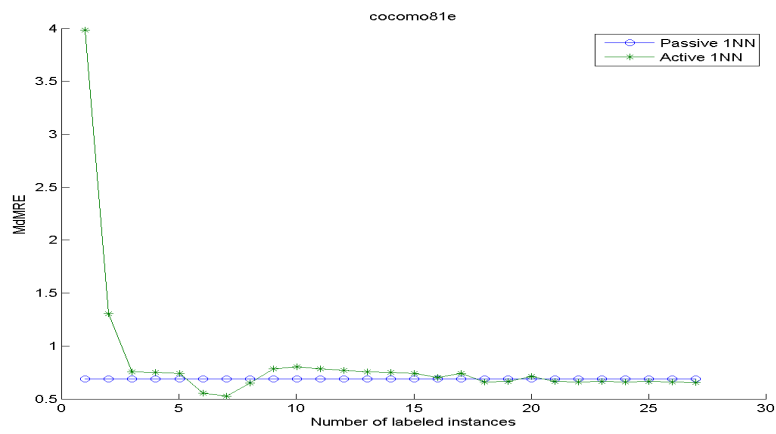
H.4 [Software Cost Estimation]: *k*-NN; D.2.8 [Software Engineering]: Cost—*active learning*



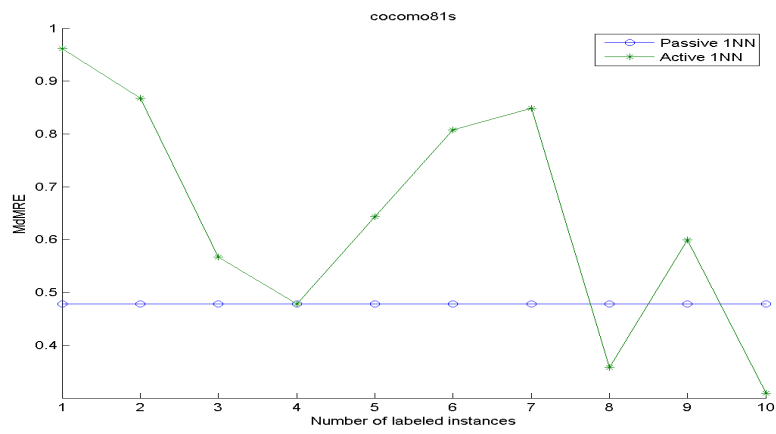
(a) cocomo81



(b) cocomo81o

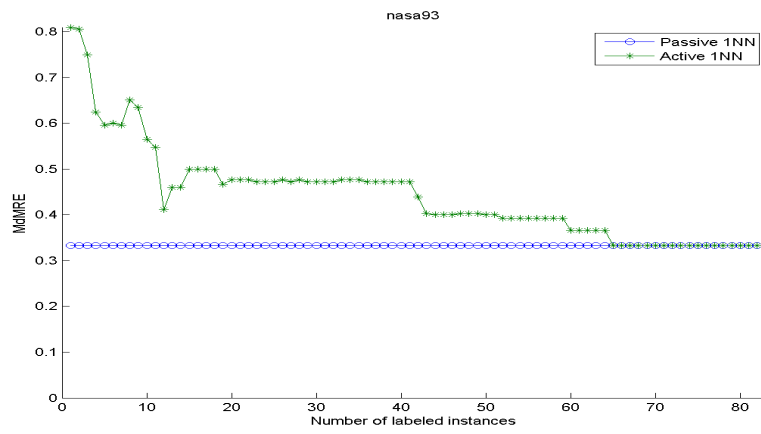


(c) cocomo81e

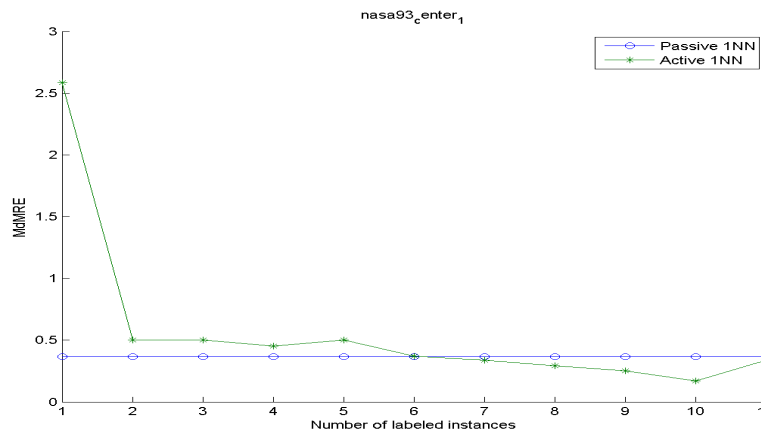


(d) cocomo81s

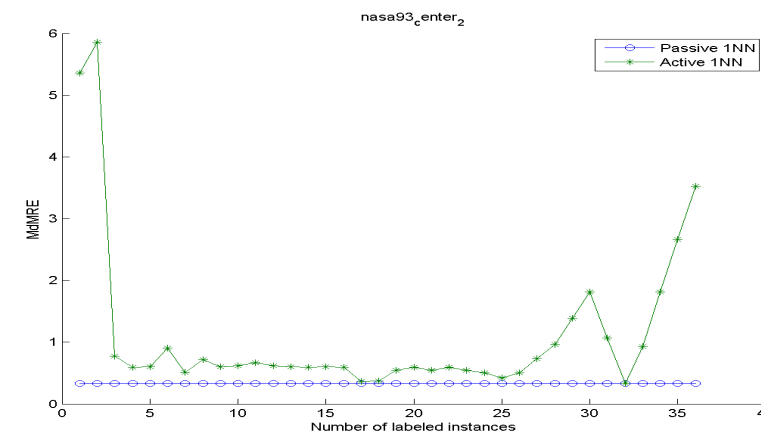
Figure 2: cocomo MdMMRE Values.



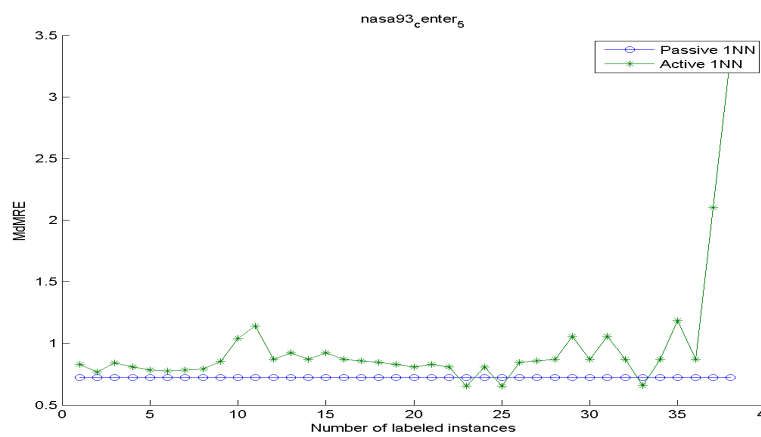
(a) nasa93



(b) nasa93_center_1



(c) nasa93_center_2



(d) nasa93_center_5

Figure 3: Nasa93 MdMMRE Values.