EvoImp: Multiple Imputation of Multi-label Classification data with a genetic algorithm

Antonio Fernando Lavareda Jacob Junior, Fabrício Almeida do Carmo, Ádamo Lima de Santana, Ewaldo Eder Carvalho Santana, Fábio Manoel França Lobato*

* fabio.lobato@ufopa.edu.br

Appendix S2 - Computational complexity of the proposed method

The computational complexity of EvoImp ($O(Method)$) is structured into three main processes:

- Initial population generation using Simple Imputation Methods ($O(SIM_i)$);
- Genetic operations: selection, crossover and mutation ($O(GenOp)$);
- Classification to measure the performance of the individuals ($O(Class_j)$).

Therefore, the complexity of EvoImp can be described by Eq. 1:

$$O(Method) = \sum_{i=1}^{n} O(SIM_i) + O(GenOp) + \sum_{j=1}^{m} O(Class_j) \quad (1)$$

where $i \in (1, \ldots, n)$ denotes the number of imputation methods adopted, and $j \in (1, \ldots, m)$ denotes the number of classification methods used for the multi-label learning tasks.

The $O(SIM_i)$ and $O(GenOp)$ complexities for this problem studied had a low impact on the equation. This fact occurs because the attributes that influence these complexity have a low value (for example, the number of individuals, the number of generations, the number of missing values, and others).

On the other hand, the complexity of most multi-label classifiers depends on the size of the database (X) and especially on the number of labels (Y) [1]. In this case, $O(SIM_i)$ and $O(GenOp)$ complexities are much smaller than $O(Class_j)$. This fact was confirmed during the computational experiments. Therefore, we can conclude that the complexity of the method is (Eq. 2):

$$O(Method) = MAX(O(Class_j)) \quad (2)$$

Among the classifiers chosen for this work, HOMER presented the worst performance, which can be justified by the balanced clustering process [2].

References

1. Bogatinovski J, Todorovski L, Dzeroski S, Kocev D. Comprehensive comparative study of multi-label classification methods. Expert Syst Appl. 2022;203:117215. DOI: 10.1016/$j.eswa$.2022.117215.
2. Banerjee A, Ghosh J. Scalable Clustering Algorithms with Balancing Constraints. Data Min Knowl Disc. 2006;13:365-395.