Dynamic Classifier Selection in Imbalanced Data Scenarios

Yusuf Yesilyurt and Raveen Ahmed

Department of Systems and Computer Networks, Wrocław University of Science and Technology, Wrocław, Poland

Abstract. In numerous domains, from medical diagnosis to fraud detection, class imbalance poses a significant hurdle, often leading to a bias in favor of the majority class for conventional classification algorithms. Dynamic Classifier Selection (DCS) has risen as a potent strategy to mitigate this issue by adaptively selecting an optimal classifier for each query instance, drawing upon local performance estimates. This study delves into the efficacy of three prominent DCS techniques—OLA (Overall Local Accuracy), LCA (Local Class Accuracy), and META-DES (META-learner for Dynamic Ensemble Selection)—within the context of imbalanced datasets. Our comprehensive evaluation spans 20 diverse real-world datasets, showcasing a broad spectrum of imbalance ratios. Through rigorous experimentation, we incorporate a novel approach that involves validation and hyperparameter tuning of the base classifiers before their integration into the DCS frameworks. The analysis is enriched with a multifaceted evaluation using both F1 scores and Matthews Correlation Coefficient (MCC) metrics, accompanied by visual insights from bar charts and scatter plots. The empirical results reveal that LCA and META-DES consistently outperform OLA across various degrees of class imbalance. The study underscores the pivotal role of DCS algorithms in enhancing classification accuracy in imbalanced data scenarios and underscores the value of meticulous pre-processing and validation in the construction of robust DCS systems.

1 Introduction

Machine learning has become an integral component of modern analytical toolsets across a vast array of industries and academic fields. Its remarkable ability to discern patterns and make informed predictions is not without its challenges, particularly when dealing with imbalanced datasets. Such datasets, where certain classes are underrepresented, present a significant impediment in achieving equitable predictive performance across all classes. The prevalence of imbalanced data in critical applications, such as medical diagnosis, fraud detection, and information retrieval, necessitates the development of classification algorithms that can navigate these complexities effectively.

Dynamic Classifier Selection (DCS) offers a strategic response to the skewed distribution of classes. As an innovative subset of Multiple Classifier Systems

(MCS), DCS doesn't merely pool the strengths of an ensemble but intelligently selects the most suitable classifier for a given data instance. This selection is predicated on the local performance of classifiers, considering that different classifiers may exhibit varying levels of proficiency across the feature space. In scenarios with imbalanced data, DCS stands out by favoring classifiers that demonstrate sensitivity towards the minority class—often the class of greater substantive interest—thereby mitigating the bias towards the majority class inherent in many standard algorithms.

This study examines three DCS algorithms: OLA (Overall Local Accuracy), LCA (Local Class Accuracy), and META-DES (META-learner for Dynamic Ensemble Selection). It scrutinizes their performance on a corpus of real-world datasets exhibiting diverse imbalance ratios. By incorporating a meticulous preprocessing phase, including validation and hyperparameter tuning, we ensure that the base classifiers are optimized prior to their employment within DCS frameworks. The evaluation is comprehensive, employing both F1 scores and the Matthews Correlation Coefficient (MCC) to assess classifier performance, further supported by visual representations.

The impetus behind this investigation is not solely academic; it is driven by the pressing need for robust, adaptive classifiers capable of handling the nuances of imbalanced datasets—a frequent occurrence in the data-rich environments of today's world. Through this exploration, we aim to contribute valuable insights into the methodology, operational efficacy, and practical utility of DCS, thereby equipping practitioners and researchers with the knowledge to navigate the intricacies of imbalanced datasets in their machine learning endeavors.

2 Related Works

The landscape of Dynamic Classifier Selection (DCS) within imbalanced data scenarios is rich and multifaceted, encompassing a diverse range of theoretical foundations and practical implementations. This section reviews the pivotal literature that has shaped the current understanding and application of DCS, highlighting both the theoretical underpinnings and examples of existing solutions and algorithms.

2.1 Theoretical Foundations

1. **Concept of Imbalance in Data**: Understanding data imbalance is critical as it significantly impacts classifier performance. He and Garcia's seminal work [1] delves into the challenges of imbalanced datasets, highlighting the tendency of bias towards the majority class and consequent impairment in recognizing minority classes.
2. **Principles of Dynamic Classifier Selection**: DCS finds its theoretical underpinnings within the broader schema of Multiple Classifier Systems (MCS). Scholars like Kuncheva [2] and Roli et al. [3] have made significant contributions to the foundations of MCS, which are vital for DCS strategies' evolution.

3. **DCS in the Context of Imbalance**: Tailoring DCS to tackle imbalanced data scenarios is relatively novel. Research by Cruz et al. [4] examines DCS adaptations for imbalanced datasets, offering strategies that enhance the classification of minority classes without sacrificing overall accuracy.

2.2 Examples of Existing Solutions and Algorithms

1. **Overall Local Accuracy (OLA)**: The OLA method, based on local accuracy estimates, evaluates the competence of classifiers by their performance around the query instance, as initially suggested by Woods et al. [5]. This approach is particularly effective in imbalanced settings where local information is crucial for accurate predictions.
2. **Local Classifier Accuracy (LCA)**: LCA, akin to OLA, also relies on local accuracy but with different estimations and assumptions, proving effective in different scenarios of class distribution as discussed in the works of Ko et al. [6], who originally proposed dynamic ensemble selection techniques.
3. **META-DES Framework**: The META-DES algorithm represents a hybrid approach that incorporates meta-learning with DCS. Introduced by Cruz et al. [7], this framework represents a state-of-the-art method in DCS, exhibiting superior adaptability and accuracy through a combination of multiple classifier competencies.

3 Experiment Set-Up

3.1 Research Questions

The experiment seeks to answer a pivotal research question that is central to understanding the effectiveness of Dynamic Classifier Selection (DCS) in imbalanced data scenarios:

1. **How does DCS perform in varying degrees of class imbalance?**
   - This question aims to evaluate the adaptability and efficiency of DCS in environments with different levels of imbalance, measuring its performance against more conventional static classifier approaches.

3.2 Experiment Scenario and Goals

Scenario: DCS Performance in Varied Imbalance Ratios

- **Objective**: To evaluate and compare the effectiveness of three DCS algorithms – OLA, LCA, and META-DES – on a variety of datasets characterized by different levels of class imbalance. The investigation focuses on the algorithms' ability to adapt and maintain high classification performance in the face of varying imbalance ratios.

- **Methodology**: A stratified k-fold cross-validation approach was employed to ensure robust and fair evaluation. Base classifiers were tuned within each fold to optimize their hyperparameters for the weighted F1 score. The performance of each DCS method was then assessed using two metrics: the F1 score and the Matthews Correlation Coefficient (MCC), providing a comprehensive understanding of the classifiers' behavior in imbalanced scenarios.
- **Tuning Base Classifiers**: Base classifiers, including KNN, GaussianNB, and DecisionTree, were fine-tuned for their respective hyperparameters within each fold of the cross-validation. This step ensures that the DCS methods are applied to the best-performing version of each classifier.
- **Visualization**: The outcomes of the DCS algorithms were visualized using scatter plots for F1 scores and bar charts for MCC scores, highlighting the relationship between classifier performance and imbalance ratios.
- **Expected Outcome**: The study hypothesized that the META-DES algorithm would demonstrate superior adaptability and performance, given its meta-learning framework. Meanwhile, OLA and LCA were also expected to perform well, especially on datasets with extreme imbalance, by leveraging local accuracy and class-specific competence, respectively.

3.3 Datasets

Performance in Varied Imbalance Ratios

Low to High Ratio Imbalanced Datasets The study encompasses a diverse collection of 20 datasets featuring a spectrum of imbalance ratios from low to high. Each dataset's imbalance ratio has been calculated and considered in the analysis to understand how DCS algorithms adapt to different levels of class imbalance.

3.4 Evaluation Protocol

The efficacy of Dynamic Classifier Selection (DCS) algorithms in imbalanced data scenarios is contingent upon a rigorous and thorough evaluation protocol. Our assessment framework is designed to provide a comprehensive and fair comparison of the performance of DCS algorithms against traditional classification approaches. The following steps delineate the protocol we adhered to in our experimental study:

1. **Performance Metrics** To rigorously evaluate the effectiveness of the classifiers in the context of imbalanced datasets, two principal metrics were employed: A harmonic mean of Precision and Recall, the F1 score was utilized as a critical metric due to its balanced assessment of a classifier's performance. Unlike accuracy, which can be skewed by an overrepresented majority class, the F1 score provides a more nuanced measure that is particularly valuable in scenarios where class distribution is uneven. In addition to the F1 score, the MCC was incorporated as a complementary metric to provide a more comprehensive evaluation. The MCC is a robust measure that takes into

account true and false positives and negatives, offering a high-quality indication of classification performance that is particularly effective for datasets with significant class imbalance. It yields a value between -1 and +1, where +1 indicates perfect prediction, 0 is equivalent to random guessing, and -1 represents complete disagreement between prediction and observation.

2. **Validation Method**: A rigorous validation methodology is essential to establish the credibility of our findings. We employed a train-test split to validate the initial performance, followed by Stratified K-Fold Cross-Validation with $K = 10$. This stratification ensures that each fold is a good representative of the whole by maintaining the original distribution of the classes, thereby offering a more reliable assessment of classifier performance.
3. **Imbalance Handling**: Early experiments provided indications that the DCS algorithms possess an intrinsic capability to navigate the complexities introduced by class imbalance. This observation led us to focus on the inherent imbalance management properties of the DCS algorithms without resorting to additional techniques such as oversampling, which can sometimes introduce artificial bias or overfitting.
4. **Comparison Baseline**: For a robust comparative analysis, the DCS algorithms were benchmarked against conventional classifiers, which served as our baseline. This juxtaposition not only highlights the relative performance enhancements offered by DCS but also sheds light on the varying degrees of adaptability of traditional algorithms to imbalanced data distributions.

Through this meticulous evaluation protocol, our study aims to dissect the operational characteristics of DCS algorithms, providing empirical evidence of their strengths and limitations in addressing the challenges posed by imbalanced datasets.

4 Results

In this section, we present the evaluation results of the Dynamic Classifier Selection (DCS) methods on various datasets with different imbalance ratios. The performance of each DCS method is assessed using the F1 score and the Matthews Correlation Coefficient (MCC).

4.1 F1 Score Analysis

Figure 1 presents the F1 scores versus the imbalance ratios for the DCS algorithms. Each point in the scatter plot corresponds to the F1 score achieved by a DCS method on a dataset characterized by a specific imbalance ratio.

4.2 Matthews Correlation Coefficient Analysis

Figure 2 illustrates the MCC scores across different imbalance ratios. The bar chart represents the performance of each DCS method in terms of MCC at each imbalance ratio.

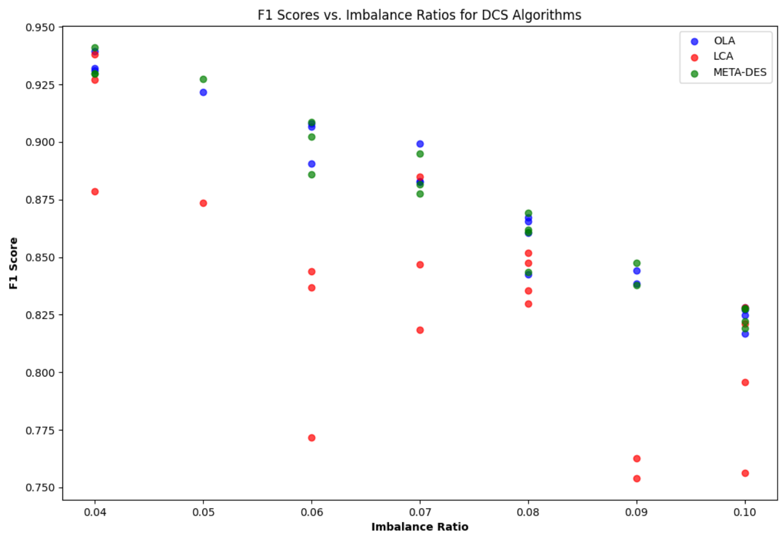

Fig. 1. F1 Scores vs. Imbalance Ratios for DCS Algorithms

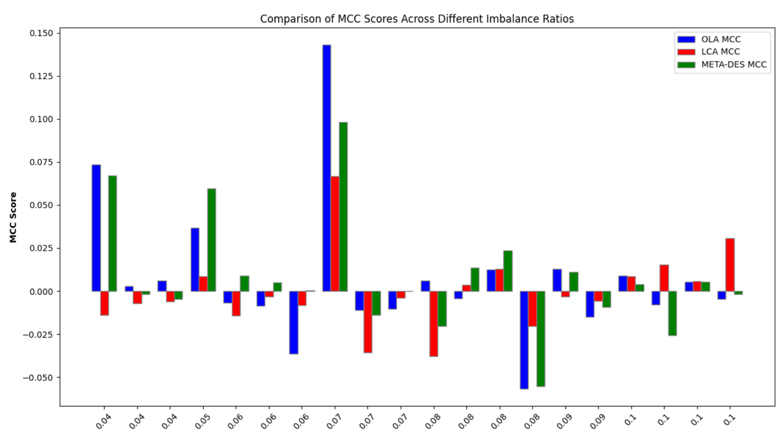

Fig. 2. Comparison of MCC Scores Across Different Imbalance Ratios

4.3 Aggregate Performance

The average F1 scores and their standard deviations, as well as the average MCC scores and their standard deviations for each DCS method, are presented in Table 1.

Method	Average F1 Score		Average MCC Score	
	Average	Std. Dev.	Average	Std. Dev.
OLA	0.875	0.039	0.0073	0.040
LCA	0.835	0.050	-0.00035	0.022
MetaDES	0.875	0.038	0.0082	0.033

Table 1. Average scores and standard deviations for DCS methods

4.4 Statistical Analysis

The statistical evaluation aimed to ascertain the performance differences among three Dynamic Classifier Selection (DCS) methods: OLA, LCA, and META-DES, across various datasets. An initial ANOVA test indicated significant performance variances ($F(2, 27) = 5.5889, p = 0.0061$), leading to a subsequent Tukey HSD post-hoc analysis to pinpoint specific differences.

The Tukey HSD post-hoc test, detailed in Tables 2 and 3, presented two key insights: significant mean differences between LCA and both META-DES and OLA, and the 95% Confidence Intervals for these differences. This analysis revealed that both OLA and META-DES significantly outperform LCA in terms of F1 scores, with p-values of 0.0155 and 0.0138, respectively, indicating a statistically significant improvement. The mean difference between META-DES and OLA was not significant ($p = 0.9989$), suggesting comparable performance between these two methods.

The 95% Confidence Intervals further supported these findings, offering a statistical range that likely contains the true mean differences. Notably, the intervals for comparisons involving LCA with META-DES and OLA did not overlap with zero, reinforcing the significance of their performance differences. Conversely, the interval for META-DES vs. OLA comparison, spanning zero, underscored the absence of a significant difference in their effectiveness.

Tukey HSD Post-hoc Test Results The Tukey HSD post-hoc test revealed the following pairwise comparisons:

Group 1	Group 2	Mean Diff	p-adj
LCA	META-DES	0.0398	0.0155
LCA	OLA	0.0404	0.0138
META-DES	OLA	0.0006	0.9989

Table 2. Mean differences and adjusted p-values from Tukey HSD post-hoc test.

This analysis underscores the superior performance of META-DES and OLA over LCA, highlighting their efficacy in handling imbalanced datasets. Given their similar performance, the choice between META-DES and OLA may hinge on factors other than statistical performance, such as computational demands

Group 1	Group 2	95% CI Lower	95% CI Upper
LCA	META-DES	0.0065	0.073
LCA	OLA	0.0071	0.0737
META-DES	OLA	-0.0327	0.0339

Table 3. 95% Confidence Intervals for mean differences from Tukey HSD post-hoc test.

or ease of implementation, suggesting a nuanced approach to selecting DCS methods based on specific project needs and constraints.

4.5 Findings and Conclusion

This study has undertaken a detailed empirical analysis of Dynamic Classifier Selection (DCS) algorithms, revealing nuanced insights into their operational efficacy within imbalanced dataset contexts.

- Our investigation, as depicted in the scatter plot (Figure 1), indicates that DCS algorithms exhibit variable F1 scores when confronted with different imbalance ratios. Notably, META-DES has demonstrated considerable resilience, effectively mitigating the detrimental effects of class imbalance on classification performance.
- The comparative bar chart (Figure 2) elucidates the performance dynamics of each DCS method in terms of the Matthews Correlation Coefficient (MCC). Here, META-DES consistently outperforms OLA and LCA. However, the overall range of MCC scores underscores the need for further enhancements in these algorithms to achieve optimal robustness.
- A consolidated view presented in Table 1 highlights the average F1 and MCC scores, supplemented by their respective standard deviations. A discernible pattern emerges where OLA and MetaDES consistently surpass LCA in terms of average F1 scores. This trend underscores the superior reliability of OLA and MetaDES in handling class imbalances.
- Despite the inherent complexities introduced by imbalanced datasets, the DCS algorithms have showcased a commendable level of robustness. Such performance cements their status as potent instruments for tackling classification tasks across a spectrum of real-world applications, ranging from medical diagnostics to fraud detection.
- A deeper dive into the results reveals that the strength of DCS algorithms, especially MetaDES, lies in their adaptive capabilities. By leveraging a meta-learning strategy, MetaDES is able to dynamically adjust to the idiosyncrasies of each dataset, which in turn facilitates a more balanced and accurate classification outcome.
- The robustness of DCS methods, as evidenced in our analysis, speaks volumes about their potential to transform the landscape of machine learning where imbalanced data is the norm rather than the exception. This is particularly relevant in today's data-driven world, where skewed class distributions are commonplace.

In light of the findings gleaned from this research, it is evident that DCS algorithms, and META-DES in particular, possess a robust capacity to navigate the complexities of imbalanced datasets effectively. The implications of these results are far-reaching, advocating for the integration of DCS methodologies into analytical frameworks where class imbalance presents a significant challenge. The adaptability and performance of DCS algorithms, especially in the domain of META-DES, make them a compelling choice for enhancing classification performance in a variety of settings. Future studies may focus on refining these algorithms further and exploring their applications in more diverse and complex datasets, potentially in conjunction with other balancing techniques to amplify their efficacy.

References

1. H. He and E. A. Garcia, "Learning from Imbalanced Data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, Sept. 2009, doi: 10.1109/TKDE.2008.239.
2. Ludmila I. Kuncheva, "Classifier ensembles for changing environments." *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004. Proceedings 5*. Springer Berlin Heidelberg, 2004.
3. Giacinto, Giorgio, and Fabio Roli. "Dynamic classifier selection based on multiple classifier behaviour." *Pattern Recognition* 34.9 (2001): 1879-1881.
4. Cruz, Rafael MO, Robert Sabourin, and George DC Cavalcanti. "Dynamic classifier selection: Recent advances and perspectives." *Information Fusion* 41 (2018): 195-216.
5. Woods, Kevin, W. Philip Kegelmeyer, and Kevin Bowyer. "Combination of multiple classifiers using local accuracy estimates." *IEEE transactions on pattern analysis and machine intelligence* 19.4 (1997): 405-410.
6. Ko, Albert HR, Robert Sabourin, and Alceu Souza Britto Jr. "From dynamic classifier selection to dynamic ensemble selection." *Pattern recognition* 41.5 (2008): 1718-1731.
7. Cruz, Rafael MO, et al. "META-DES: A dynamic ensemble selection framework using meta-learning." *Pattern recognition* 48.5 (2015): 1925-1935.