Distance Metric Recommendation for k-Means Clustering: A Meta-Learning Approach

Mark Edward M. Gonzales*, Lorene C. Uy*, Jacob Adrianne L. Sy*, Macario O. Cordel, II†
*Department of Software Technology, College of Computer Studies, De La Salle University, Manila, Philippines
†Department of Computer Technology, College of Computer Studies, De La Salle University, Manila, Philippines
{mark_gonzales, lorene_c_uy, jacob_adrianne_l_sy, macario.cordel}@dlsu.edu.ph

***Abstract*—The choice of distance metric impacts the clustering quality of centroid-based algorithms, such as k-means. Theoretical attempts to select the optimal metric entail deep domain knowledge, while experimental approaches are resource-intensive. This paper presents a meta-learning approach to automatically recommend a distance metric for k-means clustering that optimizes the Davies-Bouldin score. Three distance measures were considered: Chebyshev, Euclidean, and Manhattan. General, statistical, information-theoretic, structural, and complexity meta-features were extracted, and random forest was used to construct the meta-learning model; borderline SMOTE was applied to address class imbalance. The model registered an accuracy of 70.59%. Employing Shapley additive explanations, it was found that the mean of the sparsity of the attributes has the highest meta-feature importance. Feeding only the top 25 most important meta-features increased the accuracy to 71.57%. The main contribution of this paper is twofold: the construction of a meta-learning model for distance metric recommendation and a fine-grained analysis of the importance and effects of the meta-features on the model's output. The code and dataset are available at https://github.com/memgonzales/meta-learning-clustering.**

***Index Terms*—meta-learning, meta-features, *k*-means, clustering, distance metric, random forest**

I. Introduction

A problem with the current data mining pipeline is that each stage involves a selection from a wide array of alternatives, such as the choice of the clustering algorithm [1] and the number of clusters [2]. In the specific case of centroid-based clustering algorithms, such as k-means, another important consideration is the distance measure or metric, as it is pivotal in determining the cluster assignment of a data point. The impact of the distance metric on the overall clustering quality has been investigated in several studies [3]–[6], with Gupta and Chandra [7] reporting significant differences in the resulting accuracy and cluster separation depending on the distance metric used for k-means clustering.

Although the optimal distance measure may be selected through theoretical or experimental methods [8], the former hinges on deep domain expertise and understanding of the geometry of the dataset, while the latter demands a significant amount of time and resources. An approach to address this is via meta-learning. Derived from the idea of "learning to learn," meta-learning is a machine learning subfield that explores the automatic recommendation of algorithms and parameters [9].

In the context of clustering, Zhu et al. [8] performed an initial study on applying meta-learning to distance metric selection. However, their work considered a limited set of general, distance-based, statistical, structural, and information-theoretic meta-features, and did not provide fine-grained insights into how these meta-features influence their model's prediction of the optimal distance metric.

Our study explores a wider set of meta-features combined from other works [10], [11]. In particular, it investigates the use of *(i)* general, *(ii)* statistical, *(iii)* information-theoretic, *(iv)* structural, and *(v)* complexity meta-features in building a random forest model that automatically recommends a distance metric for k-means clustering. Additionally, it attempts to provide a fine-grained analysis of the importance and effects of the meta-features on the meta-learning model's output.

II. Related Works

The similarity measure in clustering algorithms is typically based on different distance metrics. Giancarlo et al. [12] showed that, for k-means, average link, complete link, and minimum spanning tree, the cluster qualities varied based on the distance measure. Clustering two small datasets, Bora and Gupta [5] found that correlation distance yielded the best cluster quality, albeit at the cost of a slower runtime. Their works suggest that the distance metric is critical to the cluster quality and, consequently, to the insights derived from data mining. However, in selecting the optimal choice, these studies rely on performing a computationally expensive grid search.

To avoid this resource-intensive grid searching, the application of meta-learning has been explored. Jilling and Alvarez [1] developed a clustering algorithm recommendation system using distance-based and statistical meta-features. Pimentel and De Carvalho [2] employed meta-features based on distance, evaluation, correlation, and density to determine the optimal number of clusters. Zhu et al. [8] extracted general, distance-based, statistical, structural, and information-theoretic meta-features to recommend a distance metric. While these studies establish meta-learning as a promising approach, they do not probe into how the attributes of a dataset influence the model's outputs; their meta-feature set may also be augmented.

Our work investigates a more comprehensive set of meta-features, including complexity and statistical meta-features. Our fine-grained analysis of their influence on our model's output shows that some of the additional statistical meta-features considered in our study (e.g., the mean of the sparsity) are important in predicting the optimal distance metric.

III. Dataset of Datasets

A. Dataset Collection and Preprocessing

Our methodology starts with the construction of the dataset of datasets (DoD); refer to Figure 1. There are 340 datasets collected from multiple sources: 195 from Pimentel [13], 60 from the UCI Machine Learning Repository, and 85 from Kaggle. Although clustering is unsupervised, the datasets were restricted to those with ground-truth assignments, with the number of labels used as the basis for the number of clusters. The size of our study's DoD presents an improvement over related meta-learning studies (Table I).

TABLE I
Comparison of Dataset of Datasets (DoD). The number of samples in our DoD presents an improvement over previous meta-learning studies related to clustering. Our work also considers a larger set of meta-features for the recommendation of a distance metric.

Study	Number of Dataset Entries	Number of Meta-Features	Meta-Target
[14]	57	46	Algorithm
[1]	135	25	Algorithm
[15]	219	19	Algorithm
[16]	200	19	Validation Index
[2]	219	145	Num. of Clusters
[8]	199	41	Distance Metric
Ours	340	52	Distance Metric

For each dataset in the DoD, imputation was done using either the mean (for numerical data) or mode (for categorical data). Categorical data were then converted to numerical data via one-hot encoding. The values in the datasets were normalized.

B. Meta-Target Identification

k-means clustering, coupled with a grid search over selected distance measures, was performed to label each entry in the DoD with the meta-target, i.e., the distance metric that optimizes the Davies-Bouldin score (DBS) [17]. DBS is a validation index that measures the average similarity of a cluster with its most similar cluster. Its range is $[0, +\infty)$, with lower values indicative of higher clustering quality.

Ties were broken by giving preference to the distance metric with the lowest runtime. The number of clusters was decided based on the number of ground-truth labels. Although the datasets in the collection have ground-truth assignments, clustering is an unsupervised task, and the ground-truth assignments may not be available in most real-world use cases [18]. These motivated the use of an internal validation index, specifically DBS, to evaluate clustering quality.

In identifying the meta-target, the distance measures in the work of Zhu et al. [8] were initially considered; however, this resulted in severe class imbalance. To mitigate this imbalance, the set of distance measures was restricted to only the top three metrics with the most instances, and the datasets were relabeled. In total, 139 (40.88%) datasets are classified under Chebyshev distance, 122 (35.89%) under Euclidean, and 79 (23.23%) under Manhattan.

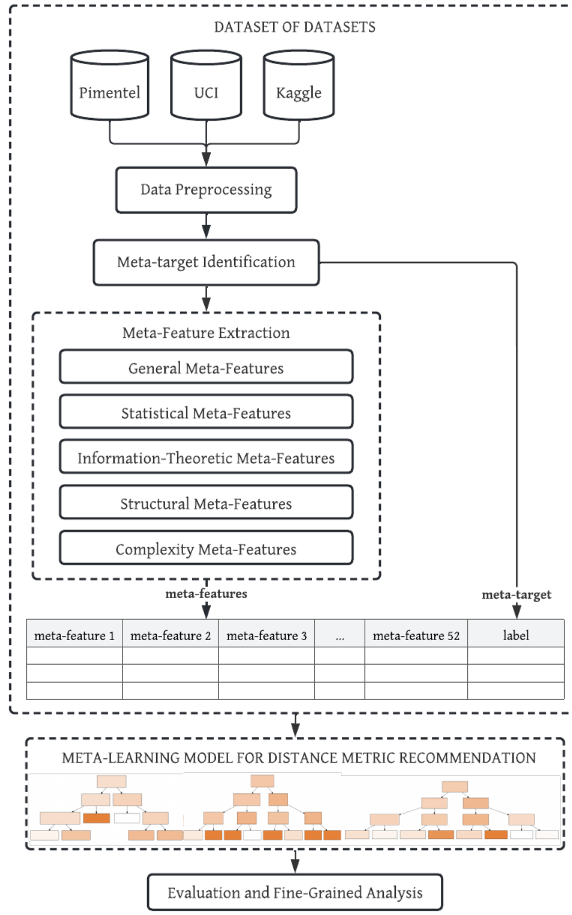

Fig. 1. Overview of the Methodology. The methodology of our study starts with the construction of the dataset of datasets, which includes dataset collection and preprocessing, meta-target identification, and meta-feature extraction. A meta-learning model for distance metric recommendation was then built, and the performance of this model was evaluated. Finally, a fine-grained analysis was conducted to determine the importance and influence of the meta-features on the model's recommended distance measure.

C. Meta-Feature Extraction

Meta-features from previous works on meta-learning [8], [10], [11] were selected based on their applicability to unsupervised tasks. In total, 52 meta-features were extracted, as enumerated in Table II; Table I compares the number of meta-features in our study with those in previous works. The meta-features can be grouped into five categories. First, general meta-features describe the dataset size and dimensionality [10]. Second, statistical meta-features cover characteristics related to feature normality, discreteness, and noisiness [10]. Third, information-theoretic meta-features quantify feature informativeness and interdependence [24]. Fourth, structural meta-features capture statistics from k-itemset frequencies [31]. Finally, complexity meta-features pertain to those related to the principal component analysis (PCA) dimensions [32].

The resulting dataset created from representing each entry in the dataset of datasets as a vector of meta-features and labeling it with the meta-target (i.e., the optimal distance metric) is hereafter referred to as the *meta-feature dataset*.

TABLE II

SUMMARY OF META-FEATURES EXTRACTED. THE DESCRIPTIONS OF THE META-FEATURES ARE LIFTED DIRECTLY FROM THE API DOCUMENTATION [19] OF PYMFE [11]. META-FEATURES MARKED WITH † INDICATE THAT BOTH THE MEAN AND STANDARD DEVIATION ACROSS ALL ATTRIBUTES WERE EXTRACTED; IN THE FIGURES HEREAFTER, THE MEAN AND STANDARD DEVIATION ARE DENOTED BY THE SUBSCRIPTS μ AND σ, RESPECTIVELY.

Category	**Number of Meta-Features**	**Meta-Features**	
		Abbreviation	**Description**
General	5	Attr-to-Inst Ratio	Ratio between the number of attributes and instances [20]
		Inst-to-Attr Ratio	Ratio between the number of instances and attributes [21]
		Num Attr	Number of attributes [22]
		Num Binary Attr	Number of binary attributes [22]
		Num Instances	Number of instances [22]
Statistical	30	Canonical Corr †	Canonical correlations of data [23]
		Correlation †	Absolute value of the correlation of distinct dataset column pairs [24]
		Covariance †	Absolute value of the covariance of distinct dataset attribute pairs [24]
		Eigenvalues †	Eigenvalues of covariance matrix from dataset [25]
		IQ Range †	Interquartile range (IQR) of each attribute [26]
		Kurtosis †	Kurtosis of each attribute [22]
		Median Abs Dev †	Median absolute deviation adjusted by a factor [25]
		Mean †	Mean value of each attribute [27]
		Median †	Median value from each attribute [27]
		Num Correlated Attr	Number of distinct highly correlated pair of attributes [28]
		Num Outliers	Number of attributes with at least one outlier value [29]
		SD †	Standard deviation of each attribute [27]
		Skewness †	Skewness for each attribute [22]
		Sparsity †	Possibly normalized sparsity for measuring the discreteness of each attribute [28]
		Trimmed Mean †	Trimmed mean of each attribute [27]
		Variance †	Variance of each attribute [24]
Information-Theoretic	4	Concentration Coeff †	Concentration coefficient of each pair of distinct attributes [30]
		Shannon's Entropy †	Shannon's entropy for each predictive attribute [22]
Structural	10	1-Itemset Min, Q1, Q2, Q3, Max	Minimum; 1st, 2nd, and 3rd quartiles; and maximum of one-itemset meta-feature [31]
		2-Itemset Min, Q1, Q2, Q3, Max	Minimum; 1st, 2nd, and 3rd quartiles; and maximum of two-itemset meta-feature [31]
Complexity	3	Ave Feat per PCA Dim	Average number of features per PCA dimension [32]
		Ave Num PCA Dim per Point	Average number of PCA dimensions per points [32]
		PCA-to-Orig Dim Ratio	Ratio of the PCA dimension to the original dimension [32]

IV. META-LEARNING MODEL FOR DISTANCE METRIC RECOMMENDATION

Framing the recommendation of the optimal distance metric as a multiclass classification problem posits the vector of meta-features as the input and the distance metric as the output. The meta-feature dataset was subjected to a 70%-30% stratified train-test split. The meta-features were fed to a random forest classifier (hereafter referred to as the *meta-learning model*), which is robust to overfitting and known to perform well on small datasets.

Finally, hyperparameter tuning was conducted via grid search with five-fold stratified cross-validation to optimize the model's accuracy. Class balancing was performed within each fold. Three techniques were explored to this end: synthetic minority oversampling technique (SMOTE), borderline SMOTE, and adaptive synthetic algorithm (ADASYN).

V. RESULTS AND ANALYSIS

A. Model Evaluation

The models were evaluated based on their accuracy (micro-averaged F1) and macro-averaged F1. The performances of the built meta-learning models are compared in Table III.

The hyperparameter space for tuning the meta-learning model is as follows (the optimal hyperparameters are given in bold): number of trees (10, **50**, 100, 150), splitting criterion (**Gini**, information entropy), maximum depth (5, **15**, 25, 25), minimum number of samples to be a leaf node (1, 2, **3**, 4), minimum number of samples to split an internal node (1, **2**, 3, 4), number of features to consider at each split ($\mathbf{log_2}$, square root), warm start (**True**, False), minimum impurity decrease (**0.0**, 0.5, 1.0), and complexity parameter α for minimal cost-complexity pruning (**0.0**, 0.5, 1.0).

TABLE III

PERFORMANCE OF THE META-LEARNING MODELS. USING BORDERLINE SMOTE IN ADDRESSING CLASS IMBALANCE YIELDED THE HIGHEST ACCURACY AND MACRO-F1 AT 70.59% AND 67.86%, RESPECTIVELY.

	SMOTE	**Borderline SMOTE**	ADASYN
Accuracy (Micro-F1)	63.73%	**70.59%**	65.69%
Macro-F1	60.29%	**67.86%**	63.01%

B. Feature Importance

To analyze the importance of the different meta-features, Shapley Additive Explanations (SHAP) [33], a game-theoretic and model-agnostic technique for interpreting an ensemble model, was utilized. Figure 2 plots the top meta-features in decreasing order of global feature importance. The top five meta-features are the mean of the sparsity, number of binary attributes, mean of Shannon's entropy, standard deviation of the variance, and standard deviation of the eigenvalues. These meta-features, with the exception of the mean of the entropy, have not been considered in the previous similar study [8].

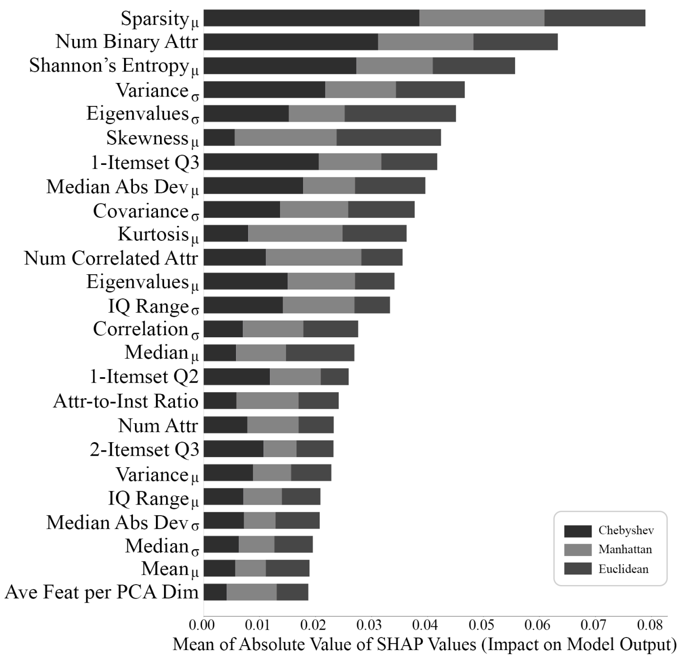

Fig. 2. Top 25 Meta-Features with the Highest Global Feature Importance. The length of each color-coded segment corresponds to the absolute magnitude of the feature's importance relative to each class. For instance, the mean of the skewness does not contribute to the prediction of Chebyshev distance as much as it does for the other two distance measures.

A fine-grained analysis was performed by limiting the domain of the computation of the SHAP values to a specific class (distance metric), as seen in Figure 3. The five most important meta-features for predicting Chebyshev distance are the mean of sparsity, number of binary attributes, mean of Shannon's entropy, standard deviation of the variance, and third quartile of the one-itemset frequencies. For Euclidean distance, these are the standard deviation of the eigenvalues, mean of the skewness, mean of the sparsity, number of binary attributes, and mean of Shannon's entropy. Lastly, for Manhattan distance, these are the mean of the sparsity, mean of the skewness, number of binary attributes, number of correlated attributes, and mean of the kurtosis.

Ergo, aside from their high global contribution, the mean of the sparsity and the number of binary attributes are also consistently among the top five meta-features with the highest importance relative to each of the three distance measures.

C. Feature Effects

The beeswarm visualizations in Figure 3 provide preliminary insights into the relationship between the values of the meta-features and their contribution to the likelihood of a prediction. The SHAP dependence plots further suggest that higher values of the mean of the sparsity, standard deviation of the standard deviation, and third quartile of the two-itemset frequencies (Figures 4a to 4d) relate to a higher probability of the model classifying an instance under Chebyshev distance. On the contrary, higher values of the mean of Shannon's entropy contribute negatively to this probability.

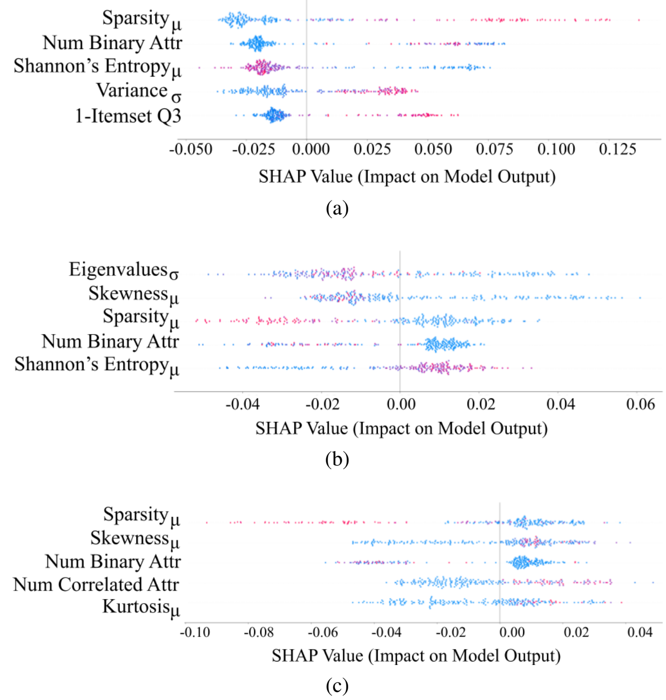

Fig. 3. Top 5 Meta-Features with the Highest Feature Importance Relative to the Prediction of (a) Chebyshev, (b) Euclidean, and (c) Manhattan Distances. Each point corresponds to the SHAP value of a meta-feature for an instance. Overlapping points are jittered vertically, providing some insight into the distribution of the SHAP values. The color indicates the value of the meta-feature; blue corresponds to lower values, while red corresponds to higher values.

Figures 4e and 4f show that higher values of the third quartile of the one-itemset frequencies and the third quartile of the two-itemset frequencies are associated with a decrease in the likelihood of classifying an instance under Euclidean distance. The same negative relationship can be observed between higher values of the mean of the eigenvalues and the likelihood of an instance being classified under Manhattan distance (Figure 4g). Meanwhile, a positive relationship exists between higher values of the mean of the canonical correlations and the said prediction probability (Figure 4h).

D. Feature Selection

The meta-features were ranked based on their global feature importance (i.e., the mean of the absolute value of the SHAP values), and the number of features was decremented by removing the least important features. The highest mean validation accuracy after performing five-fold stratified cross-validation was achieved by feeding only the top 25 meta-features (enumerated in Figure 2) to the built meta-learning model, corresponding to almost half of the entire meta-feature set. Its performance on the test set is reported in Table IV.

As seen in the confusion matrices (Figure 5), most of the misclassifications were instances under Manhattan distance that were incorrectly classified under Euclidean distance. While borderline SMOTE was applied in an attempt to address the problem of class imbalance, this result may be reflective of the underrepresentation of Manhattan distance in the dataset.

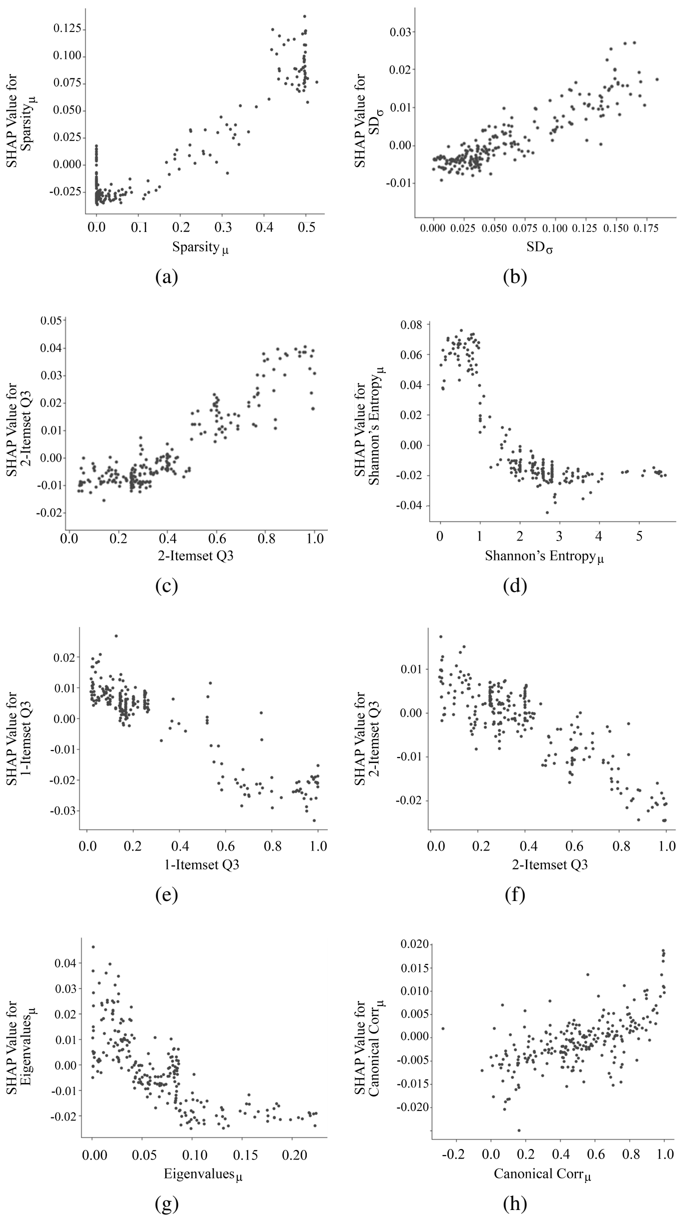

Fig. 4. SHAP Dependence Plots of (a) Mean of the Sparsity, (b) Standard Deviation of the Standard Deviation, (c) Third Quartile of the Two-Itemset Frequencies, and (d) Mean of Shannon's Entropy for Predicting Chebyshev Distance; (e) Third Quartile of the One-Itemset Frequencies and (f) Third Quartile of the Two-Itemset Frequencies for Predicting Euclidean Distance; and (g) Mean of the Eigenvalues and (h) Mean of the Canonical Correlations for Predicting Manhattan Distance. The values of these meta-features exhibit a clear positive or negative influence on the likelihood of a prediction under a particular distance metric.

TABLE IV
PERFORMANCE OF THE META-LEARNING MODEL AFTER FEATURE SELECTION. FEEDING ONLY THE TOP 25 META-FEATURES WITH THE HIGHEST GLOBAL IMPORTANCE (I.E., THE MEAN OF THE ABSOLUTE VALUE OF THE SHAP VALUES) INCREASED THE ACCURACY BY 0.98% AND THE MACRO-F1 BY 0.20%.

	All 52 Meta-Features	**Top 25 Meta-Features**
Accuracy (Micro-F1)	70.59%	**71.57%**
Macro-F1	67.86%	**68.06%**

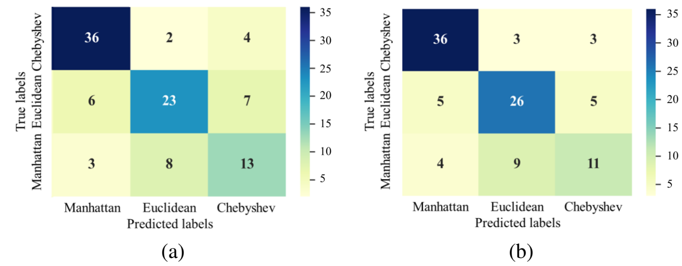

Fig. 5. Confusion Matrices of the Meta-Learning Models (a) using all the 52 meta-features and (b) using only the top 25 meta-features. The per-class F1 scores after feeding all the meta-features are 82.76%, 66.67%, and 54.17% for Chebyshev, Euclidean, and Manhattan distances, respectively. Feeding only the top 25 meta-features returned per-class F1 scores of 82.76%, 70.27%, and 51.16% for these three distance metrics, respectively.

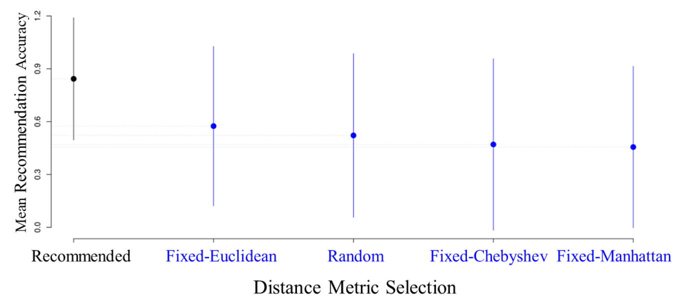

Fig. 6. Results of the Scott-Knott Effect Size Difference Test. Lines with the same color indicate that the difference between the mean RA values is negligible; inversely, lines with different colors indicate that the difference is statistically significant. The dot marks the value of the mean RA, and the length of each line is inversely related to the stability of the distance metric selection method. Aside from having the highest mean RA, our meta-learning model also has the lowest standard error of the mean (SEM) at 0.0344. For comparison, the SEM values of defaulting to Euclidean distance, defaulting to Manhattan distance, selecting a random metric, and defaulting to Chebyshev distance are 0.0448, 0.0454, 0.0461, and 0.0482, respectively.

E. Hypothesis Testing

To further evaluate the performance of the built meta-learning model, its mean recommendation accuracy (RA) [8] was compared with the mean RA values if fixed and randomly chosen distance measures were to be selected. Formally, given a dataset, let DBS_{rec} refer to the DBS if the distance metric recommended by the built model is selected and DBS_{best} and DBS_{worst} refer to the DBS values if the best- and worst-performing distance metrics are selected. The recommendation accuracy RA on this dataset is given by Equation 1.

$$RA = \frac{DBS_{\text{rec}} - DBS_{\text{worst}}}{DBS_{\text{best}} - DBS_{\text{worst}}} \tag{1}$$

As reported in Figure 6, the built model registered the highest mean RA at 83.60%, followed by defaulting to Euclidean distance at 57.46%, choosing a random distance measure at 52.15%, defaulting to Chebyshev distance at 47.02%, and defaulting to Manhattan distance at 45.56%.

Applying the Scott-Knott effect size difference test [34] showed that the mean RA of using the meta-learning model is significantly different from the mean RA values of utilizing fixed or random distance metric selection methods. These results may be taken as indicative of the effectiveness of the built meta-learning model.

VI. Conclusion

This study explored the use of a meta-learning approach for the automatic recommendation of a distance metric for k-means clustering that optimizes the Davies-Bouldin score. Five categories of meta-features were considered: general, statistical, information-theoretic, structural, and complexity. The built model yielded an accuracy of 70.59% in predicting the most optimal among Chebyshev, Euclidean, and Manhattan distance metrics for k-means clustering.

The fine-grained analysis using SHAP showed that the mean of the sparsity registered the highest feature importance globally. Limiting the meta-feature set to only the top 25 most important meta-features resulted in a slight increase in performance, bringing the model's accuracy to 71.57%. While the prediction of the minority class (Manhattan distance) posed some difficulty to the meta-learning model despite the application of borderline SMOTE, its overall mean recommendation accuracy is significantly higher compared to defaulting to fixed or randomly chosen distance measures.

References

[1] A. Jilling and M. Alvarez, "Optimizing recommendations for clustering algorithms using meta-learning," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–10.

[2] B. Pimentel and A. Carvalho, "A meta-learning approach for recommending the number of clusters for clustering algorithms," *Knowledge-Based Systems*, vol. 195, p. 105682, 02 2020.

[3] E. Xing, M. Jordan, S. J. Russell, and A. Ng, "Distance metric learning with application to clustering with side-information," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds., vol. 15. MIT Press, 2002.

[4] D. Usman and S. Sani, "Performance evaluation of similarity measures for k-means clustering algorithm," *Bayero Journal of Pure and Applied Sciences*, vol. 12, no. 2, pp. 144–148, 2020.

[5] D. Bora and D. Gupta, "Effect of different distance measures on the performance of k-means algorithm: An experimental study in MATLAB," *International Journal of Computer Science and Information Technologies*, vol. 5, 05 2014.

[6] M. O. Cordel and A. P. Azcarraga, "Fast emulation of self-organizing maps for large datasets," *Procedia Computer Science*, vol. 52, pp. 381–388, 2015, the 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).

[7] M. K. Gupta and P. Chandra, "Effects of similarity/distance metrics on k-means algorithm with respect to its applications in IoT and multimedia: A review," *Multimedia Tools and Applications*, pp. 1–26, 2021.

[8] X. Zhu, Y. Li, J. Wang, T. Zheng, and J. Fu, "Automatic recommendation of a distance measure for clustering algorithms," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 1, December 2020.

[9] C. Lemke, M. Budka, and B. Gabrys, "Metalearning: a survey of trends and technologies," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 117–130, Jun 2015.

[10] J. Vanschoren, "Meta-learning," in *Automated Machine Learning*. Springer, Cham, 2019, pp. 35–61.

[11] E. Alcobaça, F. Siqueira, A. Rivolli, L. P. F. Garcia, J. T. Oliva, and A. C. P. L. F. de Carvalho, "MFE: Towards reproducible meta-feature extraction," *Journal of Machine Learning Research*, vol. 21, no. 111, pp. 1–5, 2020. [Online]. Available: http://jmlr.org/papers/v21/19-348.html

[12] R. Giancarlo, G. Lo Bosco, and L. Pinello, "Distance functions, clustering algorithms and microarray data analysis," in *International Conference on Learning and Intelligent Optimization*. Springer, 2010, pp. 125–138.

[13] B. Pimentel, "Datasets." [Online]. Available: https://www.openml.org/s/88/data Accessed: Apr. 2, 2022.

[14] B. A. Pimentel and A. C. P. L. F. de Carvalho, "Unsupervised meta-learning for clustering algorithm recommendation," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.

[15] B. A. Pimentel and A. C. P. L. F. de Carvalho, "Statistical versus distance-based meta-features for clustering algorithm recommendation using meta-learning," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[16] S. Muravyov, A. Filchenkov, P. Brazdil, J. Vanschoren, F. Hutter, and H. Hoos, "Meta-learning system for automated clustering." in *AutoML@ PKDD/ECML*, 2017, pp. 99–101.

[17] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[18] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *2010 IEEE International Conference on Data Mining*, 2010, pp. 911–916.

[19] E. Alcobaça and F. Siqueira, "Meta-feature description table." [Online]. Available: https://pymfe.readthedocs.io Accessed: Jul. 17, 2022.

[20] A. Kalousis and T. Theoharis, "Noemon: Design, implementation and performance results of an intelligent assistant for classifier selection," *Intelligent Data Analysis*, vol. 3, no. 5, pp. 319–337, 1999.

[21] P. Kuba, P. Brazdil, C. Soares, A. Woznica *et al.*, "Exploiting sampling and meta-learning for parameter setting for support vector machines," in *Ibero-American Conference on Artificial Intelligence*, 2001, pp. 217–225.

[22] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine learning, neural and statistical classification*. Ellis Horwood Limited, 1994.

[23] A. Kalousis, "Algorithm selection via meta-learning," Ph.D. dissertation, University of Geneva, 2002.

[24] C. Castiello, G. Castellano, and A. M. Fanelli, "Meta-data: Characterization of input features for meta-learning," in *International Conference on Modeling Decisions for Artificial Intelligence*. Springer, 2005, pp. 457–468.

[25] S. Ali and K. A. Smith, "On learning algorithm selection for classification," *Applied Soft Computing*, vol. 6, no. 2, pp. 119–138, 2006.

[26] S. Ali and K. A. Smith-Miles, "A meta-learning approach to automatic kernel selection for support vector machines," *Neurocomputing*, vol. 70, no. 1-3, pp. 173–186, 2006.

[27] R. Engels and C. Theusinger, "Using a data metric for preprocessing advice for data mining applications." in *ECAI*, vol. 98. Citeseer, 1998, pp. 23–28.

[28] M. A. Salama, A. E. Hassanien, and K. Revett, "Employment of neural network and rough set in meta-learning," *Memetic Computing*, vol. 5, no. 3, pp. 165–177, 2013.

[29] C. Köpf and I. Iglezakis, "Combination of task description strategies and case base properties for meta-learning," in *Proceedings of the 2nd international workshop on integration and collaboration aspects of data mining, decision support and meta-learning*, 2002, pp. 65–76.

[30] K. Alexandros and H. Melanie, "Model selection via meta-learning: a comparative study," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 04, pp. 525–554, 2001.

[31] Q. Song, G. Wang, and C. Wang, "Automatic recommendation of classification algorithms based on data set characteristics," *Pattern recognition*, vol. 45, no. 7, pp. 2672–2689, 2012.

[32] A. Lorena, L. P. Garcia, J. Lehmann, M. de Souto, and T. Ho, "How complex is your classification problem?: A survey on measuring classification complexity," *ACM Computing Surveys*, vol. 52, pp. 1–34, 09 2019.

[33] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[34] C. Tantithamthavorn, S. McIntosh, A. E. Hassan, and K. Matsumoto, "An empirical comparison of model validation techniques for defect prediction models," *IEEE Transactions on Software Engineering*, vol. 43, no. 1, pp. 1–18, 2017.