TESTING AN ALGORITHM SELECTION MODEL BASED ON THE MAXIMUM MEAN DISCREPANCY

Yuxin Yan

Introduction

Continuous black-box optimization (BBO) problems have led to numerous reported algorithms.
Some algorithms are preferable to others given the type of problems.
Algorithm selection for solving such problems is at best cumbersome even with expert knowledge, because structures of the problems are unknown.
Algorithm selection could be handled as classification problem.
Currently Exploratory Landscape Analysis (ELA) is used for algorithm selection.
Feature engineering like developing new features is difficult and not all features have an interpretation.
We are going to test a different approach where the problem instances are described as their similarity to other problems.

Methodology

Data available to us are performances table of 24 BBOB functions, 30 instances, 2 dimensions of 5 algorithms, 1440 instances in total which could be evaluated as class label.
Exploratory Landscape Analysis (ELA) features of 1440 problem instances and distance matrix estimated using kernels with different sigma which is the band width of the kernel are used to train KNN model.

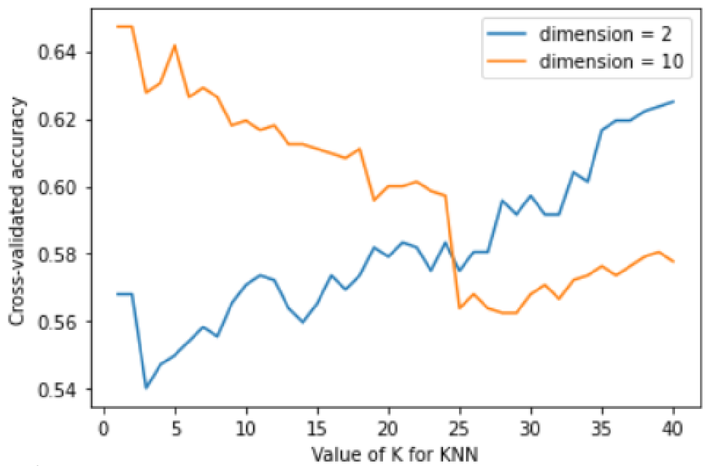


Figure 1
KNN model accuracy for k ranges from 1 to 40 for dimensionality 2 and 10 based on multi-class distance KNN (sigma=6).

Methodology

KNN is a classification algorithm that suppose we have pairs $(X_1,Y_1),(X_2,Y_2),\dots,(X_n,Y_n)$ taking values in $\mathbb{R}^d\times\{1,2\}$,where Y is the class label of X.
Given some norm $\|\cdot\|$ on $\mathbb{R}^d$ and a point $x\in\mathbb{R}^d$, let $(X_{(1)},Y_{(1)}),\dots,(X_{(n)},Y_{(n)})$ be a reordering of the training data such that $\|X_{(1)}-x\|\le\dots\le\|X_{(n)}-x\|$.see *figure 2*)

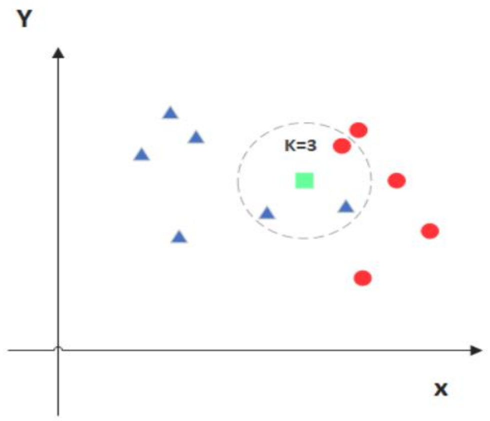
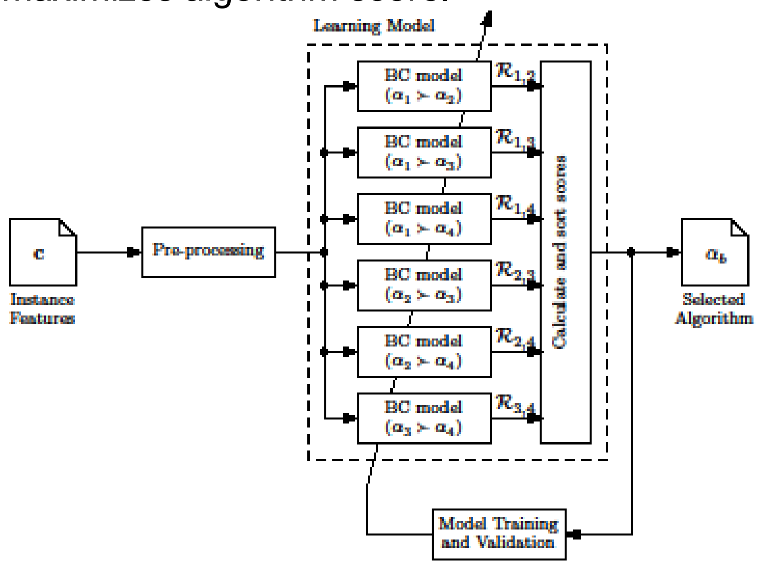


Figure 2
Calculating distance between the target data and the others, choose the k nearest points, assign its classlabel according to the most frequent classlabels among the k nearest points.

Ranking by pairwise comparison (RPC)

model generates a ranked algorithm selection preference for each problem instance (see *figure 3*). It is an ensemble of binary classification models where each model produces a probability $\mathcal{R}(\alpha_i,\alpha_j)\in[0,1]$ where $\{\alpha_i,\alpha_j\}$ are candidate algorithms.
Correspondingly, $\mathcal{R}(\alpha_j,\alpha_i)=1-\mathcal{R}(\alpha_i,\alpha_j)$.
The pairwise scores are combined to produce an algorithm score $\Sigma(\alpha_i)=\sum_{\alpha_i\neq\alpha_j}\mathcal{R}(\alpha_i,\alpha_j)$.
The selected algorithm is the one that maximizes algorithm score.



Results

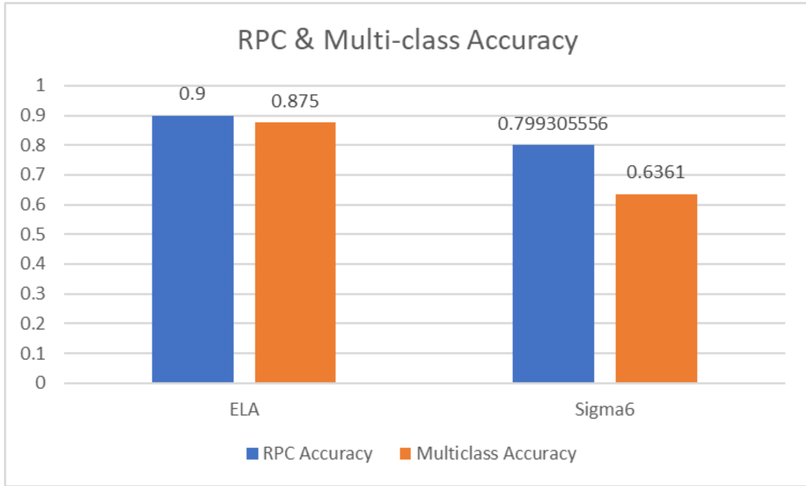


Figure 4
RPC model accuracy and multi-class model accuracy comparison based on ELA and maximum mean discrepancy calculated using sigma 6.

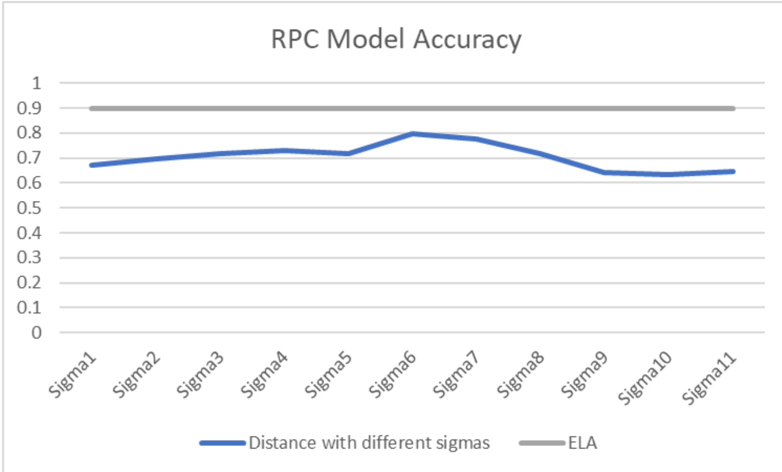


Figure 5
RPC model accuracy based on maximum mean discrepancy calculated using different sigma.

Figure 3
Structure of RPC model:
The six binary classification models take ELA as features or Distance to generate a probability score.
The total algorithm score is the summation of the corresponding probability score.

Acknowledgement

I'd like to thank the Dept. of Maths and Stats at the University of Melbourne for this opportunity. I'd also like to thank my supervisor, Andrés Muñoz-Acosta, for his guidance and many helpful conversations in this project.

Discussion

RPC model provides more accurate algorithm selection compared with multi-class classification for both ELA feature and distance matrix (see *figure* 4). It is because we have 10 pairwise models which generate an ensemble model, more models could have a more accurate prediction.
Algorithm selection model based on the maximum mean discrepancy is less accurate compared to model based on ELA features for all sigma (see *figure 5*) which could be deduced that maximum mean discrepancy captures less features of the problem instances than ELA. Also, there is no evidence showing that changing bandwidth improves the accuracy of model based on maximum mean discrepancy.
Accuracy of multi-class model based on maximum mean discrepancy increases as k increases for problems in dimension 2, while it decreases as k decreases for problems with dimensionality 10 (see *figure 1*).

Conclusion

To conclude, algorithm selection model using KNN based on maximum mean discrepancy have poorer accuracy compared to that based on ELA features. It is deduced that the distance matrix captures inappropriate data of the problem instances which are not suitable for algorithm selection. The algorithm testing set is limited, it could be possible that other algorithms perform better accuracy using the algorithm selection model. Moreover, KNN may not be the best choice to solve the BBO algorithm selection problems.

References

M.A. Mu noz and M. Kirley. (2021) Sampling Effects on AlgorithmSelection for Continuous Black-BoxOptimization. *Algorithms* 2021, 1, 0.