An Effective Feature Selection method using Monte Carlo Search

Muhammad Umar Chaudhry
Sungkyunkwan Univ..
2066, Seobu-ro, Jangan-gu,
Suwon-si, Gyeonggi-do, KOREA
+82-31-290-7987
umarch@skku.edu

Sang-Wook Kim
Hanyang Univ.
222, Wangsimni-ro, Sungdong-gu,
Seoul, KOREA
+82-2-2220-1736
wook@hanyang.ac.kr

Jee-Hyong Lee
Sungkyunkwan Univ.
2066, Seobu-ro, Jangan-gu,
Suwon-si, Gyeonggi-do, KOREA
+82-31-290-7154
john@skku.edu

ABSTRACT
Feature selection is the challenging problem in the field of machine learning. The task is to identify the optimal feature subset by eliminating the redundant and irrelevant features from the dataset. The problem becomes more complicated when dealing with high-dimensional datasets. In this paper, we propose the novel technique based on Monte Carlo Tree Search (MCTS) to find the best feature subset to classify the dataset in hand. The effectiveness and validity of the proposed method is demonstrated by experimenting on many real world datasets.

CCS Concepts

• **Information systems→Content analysis and feature selection;**

Keywords
Feature Selection; Monte Carlo Search; Heuristic Feature Selection

1. INTRODUCTION
Feature selection is one of the most complex and challenging problems in the field of machine learning [1]. It is an NP hard problem and various approaches have been developed to solve this problem. These approaches are mainly classified as Filter, Wrapper, or Hybrid approaches [2].

In this paper we use the Monte Carlo Tree Search (MCTS) [3] within the Wrapper framework to develop the heuristic approach for feature selection. We start from an empty tree node and incrementally built the tree by adding features one by one with random probability of being selected or not. The precision of tree search is then combined with generality of random sampling by the use of MCTS based search method. Classification accuracy on the current feature subset is used as a reward function and propagated back up to the root node.

RACS '17 , September 20–23, 2017, Krakow, Poland

ACM ISBN 978-1-4503-5027-3/17/09.
https://doi.org/10.1145/3129676.3130240

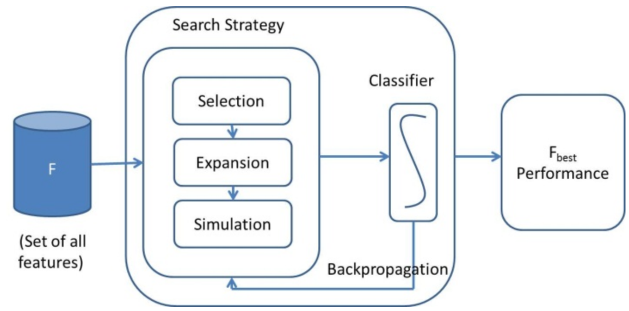

Figure 1. Proposed Method.

$$\boldsymbol{B(f_i) = Q_{j_{max}}} + \mathrm{C} \times \sqrt{\frac{\boldsymbol{2\times \ln(n)}}{\boldsymbol{n_j}}} \qquad (1)$$

Where, $Q_{j_{max}}$ is the maximum reward at the child node j, $\mathrm{C} > 0$ is a constant, n is the number of visits of current (parent) node, and n_j is the number of visits of child node j. We use the classification accuracy on chosen feature subset as a reward function for current iteration.

2. PROPOSED METHOD
.The best child at node fi is selected by slightly modifying the UCT algorithm [3], as shown in equation 1 below: The expansion at node f_i is performed by the equation 2, as shown below:

$$\boldsymbol{E(f_i) = sel(A(f_i))} \qquad (2)$$

where, $A(f_i)$ is a set of actions that can be taken at node f_i and $sel(.)$ is a function which chooses an untried action from a given set of actions.

The simulation at node f_i is performed by the equation 3, as shown below:

$$\boldsymbol{S(f_i) = rand\big(A(f_j)\big)}\ ,\ \forall\, \boldsymbol{j} \in \{\boldsymbol{i\ to\ n}\} \qquad (3)$$

where, $A(f_j)$ is a set of actions that can be taken at node f_j and $rand(.)$ is a function which chooses an action randomly from a given set of actions.

For a feature set $F = \{f_1, f_2 \ldots, f_n\}$, the set of actions at node f_i is given as:

$$A(f_i) = \{f_j, \emptyset\}$$

where, f_j and $\emptyset$ represents feature f_j is selected or not selected, respectively. The reward at current node is calculated using the classification accuracy on selected feature subset and back propagated

3. EXPERIMENTS

We conduct a series of experiments on real world datasets in order to compare and validate the effectiveness of our proposed method. The datasets are taken from LIBSVM and UCI databases.

We used the number of iterations, ranging from just 50 to 1000, as the stopping criterion, for low to high-dimensional datasets. The features subset having the maximum classification accuracy is considered as best feature subset, F_{best}. k-NN classifier with k = 1 or 5 is used as the classification algorithm with 10-fold cross validation.

As our proposed method is heuristic, therefore, we run the algorithm five times for each dataset and reported the *Best* and *Average* results of five runs in Table 1.

4. CONCLUSION & FUTURE WORK

We proposed the heuristic approach based on MCTS in wrapper framework for feature selection. The results are very promising as compared to other methods. In future, we want to investigate more on hyper parameters and to validate our approach on more datasets.

5. ACKNOWLEDGMENTS

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2017M3C4A7069440)

6. REFERENCES

[1] Guyon, I., and Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1157-1182.

[2] Brahim, A. B., and Limam, M., 2016. A hybrid feature selection method based on instance learning and cooperative subset search. *Pattern Recognition Letters.* 69 (2016), 28-34.

[3] Browne, C., Powley, E., Whitehouse, D., Lucas, S., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., and Colton, S., 2012. *IEEE Transactions on Computational Intelligence and AI in Games.* 4 (Mar. 2012).

[4] Paul, S., and Das, S., 2015. Simultaneous feature selection and weighting – An evolutionary multi-objective optimization approach. *Pattern Recognition Letters*. 65 (2015), 51-59.

[5] G, M., and Derakhshi, M. R., 2016. Feature Selection using Forest Optimization Algorithm. *Pattern recognition*. 60 (2016), 121-12

Table 1. Comparison of classification accuracy with other methods

Name of dataset	No. of features	Classifier	Proposed Method			Method 1 [4]	FSFOA [5]
			Accuracy (No. of sel features)		Std. Dev	Accuracy (No. of sel features)	Accuracy
			Best	Average			
WBDC	30	5NN	**97.72** (15)	**97.68** (17.2)	0.000702	94.06 (13.5)	---
Spambase	057	5NN	**91.72** (30)	**91.02** (32.8)	0.005159	88.48 (26)	---
Glass	009	5NN	**73.4** (6)	**72.92** (5.8)	0.002947	67.76 (4.4)	---
		1NN	**80.39** (6)	**79.29** (5.4)	0.009013	---	71.88
WBC	009	5NN	**97.57** (5)	**97.34** (7)	0.001151	96.05 (4.2)	---
Ionosphere	034	5NN	**93.73** (11)	**92.88** (13.2)	0.005895	88.31 (11.5)	89.43
Arrhythmia	195	5NN	**67.46** (103)	**66.3** (100.4)	0.006686	65.77 (100)	---
Multiple features	649	5NN	**98.5** (311)	**98.48** (321.6)	0.0004	97.88 (270)	---
Breast cancer	010	5NN	**97.66** (7)	**97.42** (7.2)	0.001747	96.53 (4.3)	---
Australian	014	5NN	**87.83** (6)	**86.84** (8.4)	0.006944	84.64 (4.7)	---
German number	024	5NN	**76.1** (10)	**75.18**(12.4)	0.0064	71.3 (10.5)	---
DNA	180	5NN	**84.65** (83)	**83.65** (86)	0.010663	83.08 (71.8)	---
Wine	013	5NN	**98.33** (6)	**98.08** (7.6)	0.00286	96.05 (6.9)	---
Vehicle	018	5NN	**74.12** (11)	**72.3** (9.2)	0.016303	65.26 (9.1)	73.98
Zoo	017	5NN	**97** (7)	**95.82** (7.6)	0.011699	95.42 (11)	---
Hillvalley	100	5NN	**59.1** (40)	**58.12** (42.4)	0.005144	57.5 (40)	---
Sonar	060	5NN	**89.4** (28)	**87.91** (27)	0.010913	82.74 (20)	86.98
Musk 1	166	5NN	**89.9** (79)	**89.35** (79.8)	0.003175	81.52 (59.3)	---
Dermatology	034	1NN	**98.09** (16)	**97.32** (16.6)	0.004703	---	97.27