Multiclass Classification of Texture Images Using Greedy Feature Selection Algorithms

Vladislav Konevsky
Samara National Research University
Samara, Russia
e-mail: konevskiyvladislav@gmail.com

Andrey Gaidel
IPSI RAS – Branch of the FSRC “Crystallography and Photonics” RAS
Samara, Russia
Samara National Research University
Samara, Russia
e-mail: andrey.gaidel@gmail.com

Rustam Paringer
Samara National Research University
Samara, Russia
e-mail: RusParinger@ssau.ru

Alexander Kupriyanov
Samara National Research University
Samara, Russia
e-mail: akupr@ssau.ru

***Abstract* — We investigated three greedy algorithms for selecting the most informative features for solving the problem of multiclass classification. The algorithms have been experimentally tested on images from the Kylberg Texture Dataset [1]. The formation of features was carried out using the MaZda software, which allows calculating the texture characteristics of the image. With the help of the algorithm of greedy forward selection, it was possible to reduce the dimension of the feature space from 298 to 141 features, and the proportion of correctly classified objects increased from 85% to 96%.**

***Keywords* — Computer science, greedy algorithms, texture analysis.**

I. Introduction

In this work, for the formation of features, the MaZda software is used, which allows obtaining 298 texture characteristics of the image. The selection of features is performed using greedy algorithms: forward selection, backward elimination, as well as a combination of forward selection and backward elimination. These algorithms make it possible to select features that are the most informative for further classification of objects into a given number of classes.

The images were obtained from the open Kylberg Texture Dataset database, which provides 576 × 576 sample images belonging to six different classes. To check the results obtained, a classification was carried out using a random forest model.

II. Description of the Dataset

As a dataset, a number of textured surfaces were used and represented in the dataset, the Kylberg Texture Dataset v. 1.0 [1]. The images are shown in figure 1. The dataset consisted of 240 images and 6 types of images.

- Canvas - Woven linen canvas.
- Cushion - Woven fabric on a cushion.
- Linseeds - Linseeds on a flat surface.
- Sand - Sand surface.
- Seat - Woven fabric on chair.
- Stone - Flat part of a granite base of a sculpture.

Each class was rendered with only one lighting setting from one direction and at the same distance. Images were captured using a Canon EOS 550d camera and a 17–70 mm Sigma lens.

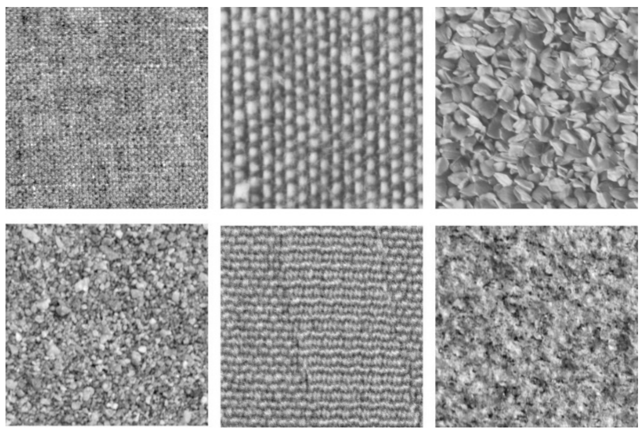

Figure 1 – Dataset

III. TEXTURE ANALYSIS

Texture analysis was carried out using the Mazda software package.

MaZda package is an efficient and reliable software tool for analyzing text images. Its effectiveness has been proven by contributors to various projects as well as other researchers who have implemented this software for a variety of texture analysis tasks. Compared to other free text analysis software (such as Keyres or LS2W), it provides a more complete image texture analysis (characterization, classification and segment) [2].

When conducting texture analysis, 298 features were obtained.

IV. DESCRIPTION OF ALGORITHMS

A. Feature selection problem.

The selection of features is carried out in order to subsequently solve the classification problem

Let Ω be the set of objects to be recognized. The set Ω was split into L classes $\Delta = \{\Omega_j\}_{j=0}^{L-1}$

To solve the classification problem, it is necessary to construct an operator $\widetilde{\Phi}(x)$ which connects x features with a class.

The probability of erroneous recognition is estimated as

$$\varepsilon = \frac{\left|\{x \in \widetilde{U} \mid \Phi(x) \neq \widetilde{\Phi}(x)\}\right|}{|\widetilde{U}|}$$

$\Phi(x) -$ ideal recognition operator

$\widetilde{U} -$ control sample

$U -$ training sample

$$\widetilde{U} \cap U = \emptyset$$

To solve the classification problem, a random forest model was used. When choosing a model, we relied on the article " Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?" [3].

The task of feature selection is to select a certain subset of features $Q \subseteq \cap\ [1;\ K] \cap Z$, ensuring the minimum classification error ε.

B. Greedy forward selection

At the first step, the feature set is empty, $Q_{(0,0)} = \emptyset$

At the *i* step, an error ε_j is sought for feature sets with the addition of a feature j

$$\tilde{Q}_{(i,j)} = Q_{(i,j-1)} \cup \{j\}$$

Added $\{j\}$ to $Q_{(i,j)}$ for which ε_j is minimal.

The algorithm stops when the required number of features is reached.

C. Greedy backward elimination

At the first step, the feature set is full features,

$$Q_{(0,0)} = Q \subseteq [1; K]$$

At step i, the error ε_j is sought for feature sets with the removal of a feature j

$$\tilde{Q}_{(i,j)} = Q_{(i,j)} \cap \{j\}$$

Removed $\{j\}$ to $Q_{(i,j)}$ for which ε_j is maximal.

The algorithm stops when the required number of features is reached.

D. Combining greedy backward elimination and forward selection

At the first step, the feature set is full features,

$$Q_{(0,0)} = Q \subseteq [1; K]$$

At step i, the error ε_j is sought for feature sets with the removal of a feature j

$$\tilde{Q}_{(i,j)} = Q_{(i,j)} \cap \{j\}$$

Removed $\{j\}$ to $Q_{(i,j)}$ for which ε_j is maximal.

After, searched $\tilde{Q}_{(i,j+1)}$

$$\tilde{Q}_{(i,j+1)} = Q_{(i,j)} \cup \{j\}$$

If the error for set $\tilde{Q}_{(i,j+1)} < Q_{(i,j)}$then the feature is added

The algorithm stops when the required number of feature or a limited number of iterations is reached.

V. Results

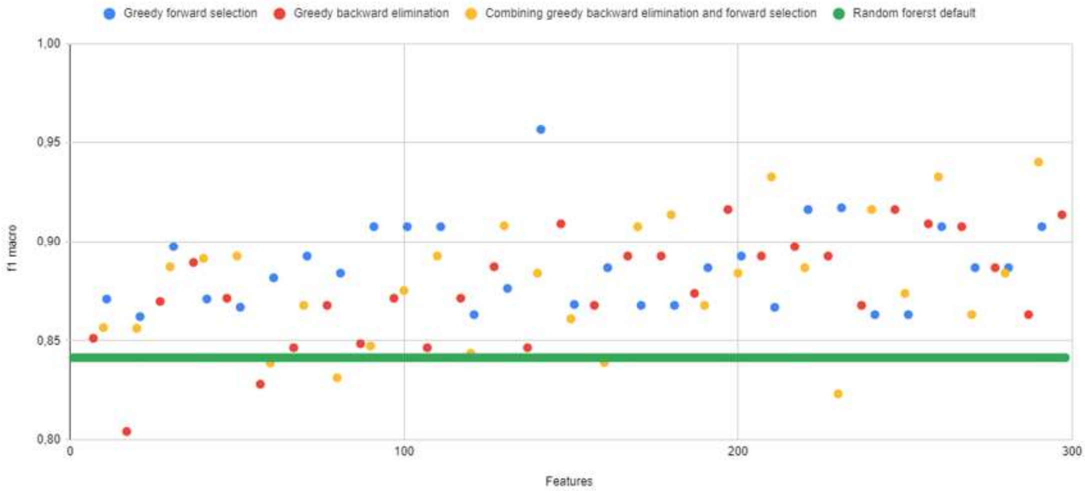

Figure 2 – F Macro

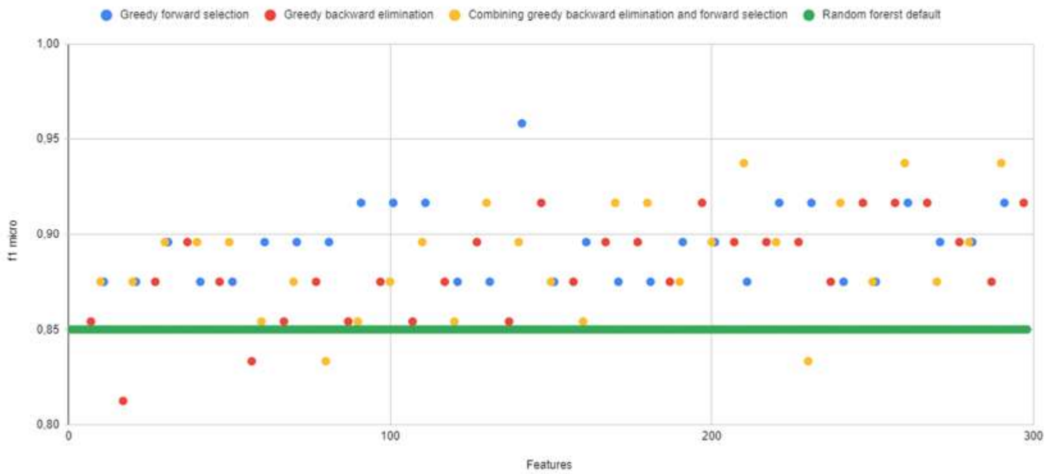

Figure 3 – F Micro

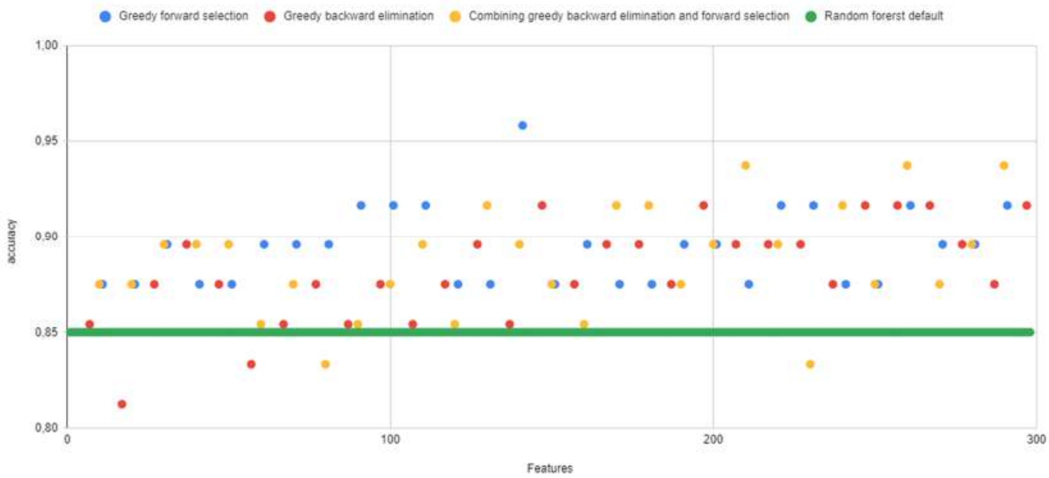

Figure 4 – Accuracy

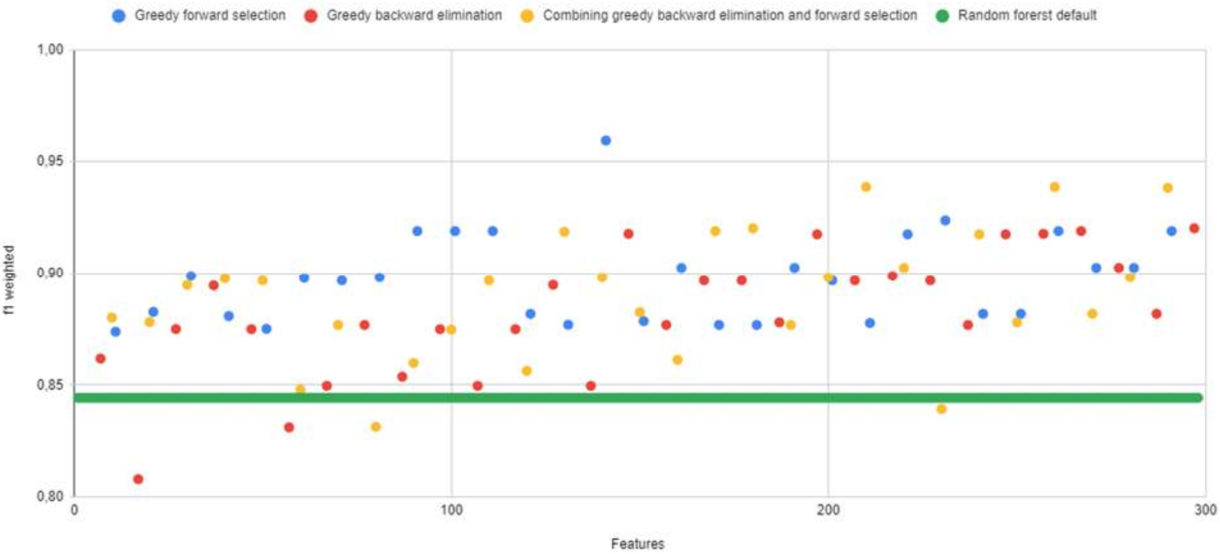

Figure 5 - Weighted

VI. Description of the Dataset

Texture analysis was carried out using the Mazda software package.

VII. Conclusion

Thus, in this work, three greedy feature selection algorithms were investigated for solving the multiclass classification problem. The algorithms have been experimentally tested on Kylberg Texture Dataset images.

The analysis of the results showed that the removal of unnecessary features leads to an improvement in the results of the classifier.

The algorithm Greedy forward selection degrades performance with a large number of functions. Whereas the Greedy backward elimination and the Combining greedy backward elimination and forward selection work better with more functions.

Combining the work of the Greedy backward elimination and Greedy forward selection algorithms increases the quality of the Greedy backward elimination algorithm, but it is inferior to the work of the Greedy forward selection algorithm.

The original feature space consisted of 298 texture characteristics generated using the MaZda software. The use of all the generated characteristics made it possible to reliably classify 85% of the images. As a result of the feature selection procedure, it was possible to reduce the dimension of the feature space to 7 values without losing the quality of the classification.

The maximum improvement in the classification quality is achieved by the Algorithm and Greedy forward selection for 141 features. Its rate is 96%.

Acknowledgment

This work was supported by the Russian Foundation for Basic Research and RA Science Committee in the frames of the joint research project RFBR 20-51-05008 Arm_a and SCS 20RF-144.

References

[1] G. Kylberg. *The Kylberg Texture Dataset v. 1.0*, Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, External report (Blue series) No. 35. Available online at: http://www.cb.uu.se/~gustaf/texture/, 2011.

[2] M. Szczypinski, M. Strzelecki, A. Materka, "*MaZda – a Software for Texture Analysis*", International Symposium on Information Technology Convergence, 6 p., DOI:10.1109/ISITC.2007.15, 2007.

[3] M. Fern´andez-Delgado, E. Cernadas, S. Barro, D. Amorim, "*Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?*", Journal of Machine Learning Research, pp. 3133-3181, 2014.