July 15, 2019

Dear Editors,

Enclosed please find our manuscript *"Eliminating the Variability of Cross-Validation Results for Some Randomized and Parallelized Learning Algorithms"* that is submitted to ReScience C as a Letter.

Recently, the reproducibility of machine learning results has become a matter of debate in the top scientific journals. These issues have also been discussed within the Artificial Intelligence and Machine Learning communities. Cross-validation is probably the most popular experimental technique for evaluating machine learning algorithms and for fine-tuning their parameters. It is natural to parallelize cross-validation by distributing different combinations of folds to different threads. Thus, it is desirable to have reproducible parallelized cross-validation results across different computing platforms.

Our submission describes a technique for improving the reproducibility of parallelized cross-validation when the underlying algorithm uses a pseudo-random number generator. This recipe is applied to LIBLINEAR, a popular library that implements randomized learning algorithms based on support vector machines. The potential audience includes users of LIBLINEAR who want to parallelize cross-validation in a way that is reproducible across different operating systems. The general recipe described in the paper, however, can be used with other libraries.

A set of patches that implement these steps so that they can be used on Linux, macOS, and Windows is a part of the submission. The repository for these patches is https://github.com/sukhoy/cvrep. It also includes a README.TXT file, which gives specific instructions for applying the patches, the original LIBLINEAR license (because the patches include snippets of its source code as context lines), and the license for our patches. In both cases, the license is the 3-clause BSD license. The only difference between the two licenses is in the copyright statement.

Both authors have read the manuscript and the text included with the patches. Both authors are aware of the current submission and consent that it will be reviewed. Both authors agree with the Open Access policy used by ReScience C.

Please do not hesitate to contact me if you have any questions or if you need any additional information about this submission.

Sincerely,

Vladimir Sukhoy
Graduate Student
Department of Electrical and Computer Engineering
1638 Howe Hall
Iowa State University
Ames, IA 50011, USA
Phone: (630) 445-1332
E-mail: sukhoy@iastate.edu
URL: http://sukhoy.public.iastate.edu/