Predicting the convergence of an iterative method from matrix images using CNN

Ryo Ota
University of Tsukuba
Tsukuba, Japan
s1821611@s.tsukuba.ac.jp

Hidehiko Hasegawa
University of Tsukuba
Tsukuba, Japan
hasegawa@slis.tsukuba.ac.jp

1 INTRODUCTION

In the field of scientific and technical computing, problems are often reduced to solving linear equations numerically. The biconjugate gradient (BiCG) method is widely used for solving linear equations with a large non-symmetric sparse matrix, but its convergence cannot be determined before trying to solve it. If the convergence were predicted in advance, we would know whether to choose a different solver.

Recently, pattern recognition in images has become much easier with the help of machine learning. To predict whether each matrix can be solved by the BiCG method, we tried to learn some relationship between matrices' grayscale images and their convergence in the BiCG method by using a convolutional neural network (CNN).

Our evaluation using 5-fold cross-validation obtained an average accuracy exceeding 80% for 28×28-pixel grayscale images.

2 EXPERIMENTS AND RESULTS

We used 875 of the 982 real non-symmetric square matrices in the SuiteSparse Matrix Collection [1]. Among the matrices used, matrix dimension ranged from 5 to 445,315 and sparsity varied from 0.000006% to 76%. Taking the logarithm base 10 of the absolute values of elements obtained values distributed from −324 to 28.

The convergence condition in the BiCG method is defined as the relative residual norm being less than or equal to 10^{-6}. Using this convergence condition, 235 matrices converged.

We used two methods, the SuiteSparse method and sigmoid method, to convert the non-zero values into the 256 grayscale values from 0 to 255. The SuiteSparse method [2] assigns a grayscale value of 128 to the median of the non-zero values, whereas the sigmoid method normalizes the non-zero values from 0 to 255. Image sizes were 28×28, 56×56, 112×112, and 224×224 pixels.

In the 28×28-pixel case, the average grayscale value generated by the SuiteSparse method for 875 matrices was 136.7 and the variance was 6567, whereas those generated by the sigmoid method were 135.0 and 2279, respectively.

We prepared two groups of datasets: dataset F, having a complete set of a convergent matrices and the same number of non-convergent matrices (470 matrices in total), and dataset G, which uses all 875 matrices. We evaluated these datasets using 5-fold cross-validation.

Table 1 shows the average accuracy. From Table 1, we see that dataset G is superior to F in all cases, with the average accuracy for dataset G exceeding 80%. The highest average accuracy was obtained in the case of 28×28 pixels, and accuracy decreased with increasing image size.

Table 2 shows the classification accuracy for dataset G for the 28×28-pixel case with the SuiteSparse method. The F-measure, which represents the performance of detecting both convergence and non-convergence correctly, is 0.81.

Table 1: Average accuracy (%)

Method	Dataset	28×28	56×56	112×112	224×224
Suite Sparse	F	80.0	78.9	79.7	78.9
	G	86.1	84.9	81.9	82.1
sigmoid	F	82.7	77.6	81.2	77.0
	G	84.1	83.3	82.6	79.3

Table 2: Classification accuracy: dataset G, 28×28 pixels, SuiteSparse method

		Real		
		Not conv.	Conv.	Total
Prediction	Not conv.	581 (66.4%)	63 (7.2%)	644
	Conv.	59 (6.7%)	172 (19.7%)	231
	Total	640	235	875

We also evaluated the matrices after standardizing. With this standardization, 289 matrices converged and 586 matrices did not. In this case of 28×28 pixels with the SuiteSparse method, the average accuracy was also over 80% and the F-measure is 0.82. The average accuracy of the SuiteSparse method was also higher than that of the sigmoid method.

A dataset of smaller images gave a higher accuracy with the SuiteSparse method, whereas all datasets gave almost the same accuracy with the sigmoid method. However, the SuiteSparse method had a higher accuracy than the sigmoid method. Based on these observations, to predict the convergence of the BiCG method for sparse matrices using grayscale images, it may be effective to perform a rough conversion of the magnitude relationship of non-zero values into grayscale values rather than a detailed conversion.

ACKNOWLEDGMENTS

This research was supported by JSPS KAKENHI Grant Number JP17K00164.

REFERENCES

[1] Tim Davis, Yifan Hu, and Scott Kolodziej. 2018. The SuiteSparse Matrix Collection. https://sparse.tamu.edu.
[2] Timothy A Davis. 2015. SuiteSparse: A suite of sparse matrix software. http://faculty.cse.tamu.edu/davis/suitesparse.html.