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

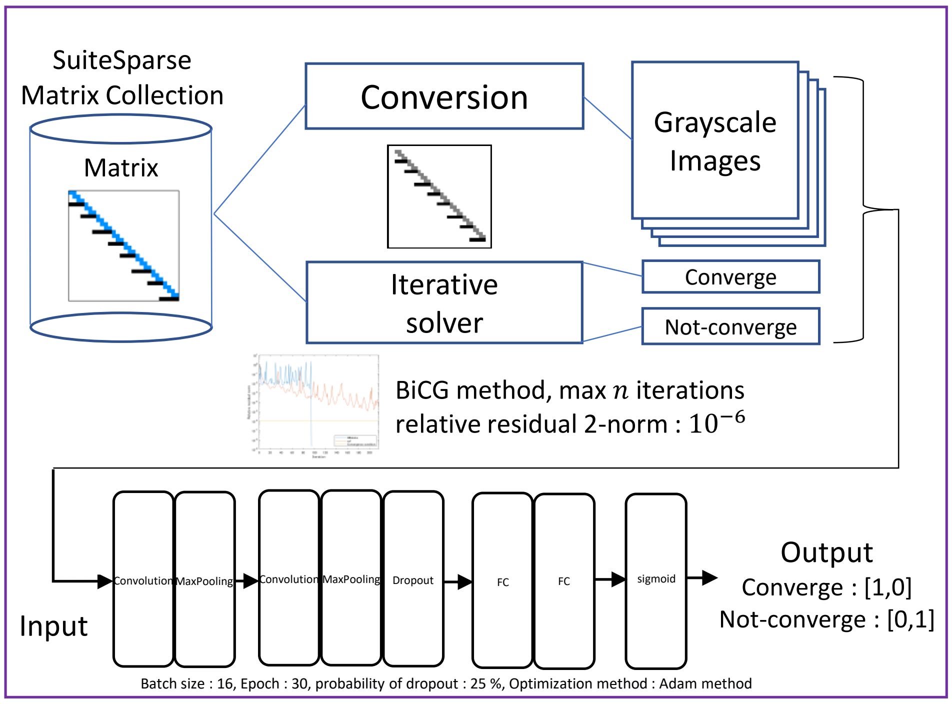
Ryo Ota and Hidehiko Hasegawa

University of Tsukuba, Japan

Mail to: s1411506@klis.tsukuba.ac.jp

Introduction

- Sometimes the iterative method converges within n (the dimension of matrix), but sometimes it does not converge at all.
- If the likelihood that the iterative method would converge could be roughly predicted in advance, a better alternative solver could be applied if that solver is unlikely to achieve convergence.
- Convolutional Neural Network (CNN) is producing remarkable progress within the field of image recognition and computer vision.
- If sparse matrices are converted to images, it would be possible to predict the convergence using CNN.
- We tested predicting the convergence of BiCG method using CNN.



Our Model

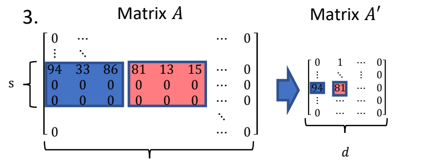
Matrix to Image

- We used 8-bit grayscale images because it is easy to classify grayscale images compared to color images.
  - Grayscale values range 0 (black) to 255 (white).
  - Image size is changed among four option 28 × 28, 56 × 56, 112 × 112, and 224 × 224 pixels.
- We used SuiteSparse [1] method and sigmoid method for their conversion.

Algorithms

1. Take absolute value of A.
2. Divide n (dimension) by d (image size) and round up. That value is named s.
3. Split the matrix into $s \times s$ blocks, get the maximum and construct matrix A'.
4. Compute the median Me, the standard deviation σ, and average μ of the logarithm base 10 of the nonzero elements of A'.
5. For $a'_{i,j}$, nonzero elements of A':

1. $A \leftarrow fabs(A)$
2. $s \leftarrow \lceil \frac{n}{d} \rceil$
3.



4. Compute Me, σ and μ for $\log_{10}(\text{nonzeros}(A'))$

SuiteSparse method

5a. Give grayscale values by following.

$$grayscale = \begin{cases} 0 & (\log_{10}(a'_{i,j}) < Me - \sigma) \\ 128 + \left\lfloor \frac{127(\log_{10}(a'_{i,j}) - Me)}{\sigma} \right\rfloor & (Me - \sigma \le \log_{10}(a'_{i,j}) \le Me + \sigma) \\ 255 & (\log_{10}(a'_{i,j}) > Me + \sigma) \end{cases}$$

This gives the grayscale value of 128 to the median of nonzero values.

sigmoid method

5b. Standardize

$$u'_{i,j} = \frac{\log_{10}(a'_{i,j}) - \mu}{\sigma}$$

and convert to grayscale values used by sigmoid function:

$$grayscale = \left\lfloor \frac{255}{1+\exp(-u'_{i,j})} \right\rfloor$$

6. If the number $\left\lceil \frac{n}{s} \right\rceil$ is less than d, then enlarge the image.

Distribution

- We use 875 of 982 non-symmetric real square matrices stored in the SuiteSparse Matrix Collection [2].
  - The matrix dimension ranges from 5 to 4453150, and sparsity varies from 0.000006 % to 76 %.
  - The absolute value of elements of A are distributed from 10^{-324} to 10^{28}.
  - After scaling, the absolute value of elements of CA are distributed from 10^{-324} to 10^{302}.

Table1. Number of matrices in each class

	Converge	Not-converge	Total
Original A	235	640	875
With scaling CA	289	586	875

$$c_{ii} := \begin{cases} 1 & (a_{ii} = 0) \\ |a_{ii}|^{-1} & (a_{ii} \neq 0) \end{cases}$$

Table2. Distribution of grayscale values in 28 × 28 pixels, except zero elements

	SuiteSparse method		sigmoid method	
	μ	σ	μ	σ
Original A	136.7	81.03	135.0	47.73
With scaling CA	136.5	79.34	135.1	47.14

- Distribution of grayscale values changed little by scaling.
- Grayscale values of SuiteSparse method distributed roughly.
- Grayscale values of sigmoid method were distributed in narrow range.

Experiments

- Dataset F consists of 235 of convergence and non-convergence matrices. Dataset G consists of all matrices (converge 235 and not-converge 640). Dataset H consists of all matrices after scaling (converge 289 and not-convergence 586).
- 5-fold cross validation was used.

Table3. Confusion Matrix

		Prediction	
		Not-converge	Converge
True	Not-converge	True Negative (TN)	False Negative (FN)
	Converge	False Positive (FP)	True Positive (TP)

$$Accuracy = \frac{TN + TP}{TN + FN + FP + TP}$$

Table4. Average accuracy by 5-fold cross validation (%)

Method	Dataset	28 × 28	56 × 56	112 × 112	224 × 224
SuiteSparse	F	80.0	78.9	79.7	78.9
	G	86.1	84.9	81.9	82.1
	H	84.0	85.3	83.9	83.4
sigmoid	F	82.7	77.6	81.2	77.0
	G	84.1	83.3	82.6	79.3
	H	81.6	83.1	83.1	81.8

1. Average accuracy achieved around 80 %.
2. Dataset G was superior to the dataset F in all options.
3. The larger the image size was, the lower the accuracy became.
4. The average accuracy of dataset G and H didn't change so much.
5. The average accuracy of SuiteSparse method and sigmoid method were about the same and mistakes were duplicated.
6. In case of convergence condition set in 10^{-10}, average accuracy achieved 80 %.

References

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