International Journal of Imaging and Robotics,
[Formerly known as the "International Journal of Imaging" (ISSN 0974-0627)]
Volume 16; Issue No. 1; Year 2016; Int. J. Imag. Robot.
ISSN 2231–525X; Copyright © 2016 [International Journal of Imaging and Robotics]

Computational Cost of Learning Vector Quantization Algorithm For Malaria Parasite Classification in Realtime Test

Iis Hamsir Ayub Wahab[1,3], Adhi Susanto[1], P. Insap Santosa[1], Maesadji Tjokronegoro[2]

[1]Department of Electrical Engineering and Information Technology,
Universitas Gadjah Mada, Yogyakarta

[2]Faculty of Medicine, Universitas Gadjah Mada,
Yogyakarta, Indonesia

[3] Department of Electrical Engineering, Universitas Khairun, Ternate, Indonesia
E-mail: [1]iishamsiraw_s3_09@mail.ugm.ac.id

ABSTRACT

Analysis and interpretation of malaria parasite images were performed in which one of them was to obtain the parasite image patterns, thus it could be conducted classification process towards image based on its pattern. The parasite image pattern is different between one and another, this depends on parasite type. Differentiating between one image and another needs a feature for each pattern. This study, therefore, aimed to analyze and evaluate algorithms of learning vector quantization neural network for malaria parasite pattern recognition test in real time. The result of this study showed that the LVQ network classification method could recognize 92% object,. Algorithm time complexity for LVQ is O(n).

Keywords: LVQ, Parasite, Realtime, Big (O)
Mathematics Subject Classification: 68T10 , 68U07
Computing Classification System: I.5

1. INTRODUCTION

Malaria is a serious disease caused by a blood parasite named Plasmodium spp. World Health Organization (WHO) in its report issued on December 2013 noted that there remained 207 malaria cases in 2012 with approximately 627 thousands mortality rate dominated by children (WHO, 2014).

The malaria definitive diagnosis was performed by finding parasite within the blood film using microscope. This is, however, a routine examination and time consuming. In addition, field study showed that the agreement level among scientists regarding this diagnosis was quite low (Arum et al., 2006).

Malaria parasite automatically detecting has been conducted by some researchers in which the number of malaria parasite in blood may be calculated based on blood sample acquisition converted into digital image. The detecting system is generally constructed through several processing stages: image acquisition, image pre-processing, image segmentation, image extraction and classification. Sio et al., (2007) developed rapid and accurate automatic system to count malaria parasite based on image analysis.

Image was obtained from acquisition result using video 3-CDD camera. There were four stages performed to be able to calculate parasitemia number, i.e.: peripheral detection, edge linking, clump linking, and parasite detector. Parasite might be identified based on peripheral properties of red blood cell by using sobel operator and Hough transformation (Le et al., 2008; Ma et al., 2010). The process employed by sobel operator was the one from a convolution that already determined toward the detected image. While Hough transformation performed by Ma et al., (2010) was to determine geometry parameter of *Plasmodium falciparum* parasite. This was also conducted by (Díaz et al., 2009), however it's on the system constructed using low pass-filter at image pre-processing, subsequently followed by extracting pre-processing image result in order to get information which red blood might have been infected by malaria parasite.

Moreover, another way to be able to detect malaria parasite from red blood cell on image might be by performing segmentation process. Segmentation was intended to separate between red blood cell and the parasite itself. Makkapati & Rao, (2011) performed color base segmentation on HSV color chamber. The image that would be segmented was first filtered using *window* 3 x 3 median filter dimension to eliminate noise. Other segmentation method used to detect malaria parasite on image was normalized cut (NCut). As another method, NCut was a non-guided segmentation method, based on global criteria.

Malaria parasite detector might be improved its performance if it used machine learning strategy (Khan et al., 2011). The constructed system was the classifying one based on typical input of each parasite type. Premaratne, Karunaweera, & Fernando, (2006) and Somsekar (2011) used back propagation neural network by histogram characteristic as input pattern to recognize malaria parasite object typed of *Plasmodium falciparum*. Histogram feature on gray scale image was also employed by Mandal, Kumar, Chatterjee, Manjunatha, & Ray, (2010) as well as Makkapati & Rao, (2011). However, the backpropagation algorithm has the disadvantage that it becomes very slow. However, backpropagation algorithm has a weakness it becomes very slow in the flat areas of the error function. To overcome these disadvantages, we need an algorithm that can detect rapidly. one of the fast classification algorithm is vector quantization. it has been successfully used in various applications (Kekre and Sarode, 2009).

Processes that do the above would require a completion algorithm. Thus, the system is built to have a good level of reliability. On the hardware specifications of the same computer, the most influence on the processing time is the number of software or programs simultaneously active at the time of testing. Therefore, the correctness of the algorithm should be tested with a certain amount of input to see the performance of the algorithm. Evaluation may include the time required to run the algorithms and memory space required for data structures (Sedgewick and Flajolet, 2013). The amount of the valuation models used to describe the time and space complexity of the algorithm.

Of these result exposures, it's still needed method development that could diagnose and accelerate reading automatically and precisely. The purpose of this study was therefore to analyze and evaluate learning vector quantization neural network for recognition of malaria parasite pattern on the real time test.

2. MATERIALS AND METHODS

In this study, sample data used was obtained from http://dpd.cdc.gov and patient blood sample data was on microscope slide glass. Analogue sample was then transformed into a digital image using monocular microscope with CMOS sensor in which its result was stored in jpg form with 24 bit color depth level. Image dimension resulted was 256 x 256 pixels. The computer specification used in this experiment was: AMD Athlon™X2 Dual Core 2.20 GHz processor, 1.74 GHz random access memory, and 250 GB storage memory.

Malaria parasite that would be tested was the falciparum type with total sample 96 consisted of:

- 24 for Gametocyte,
- 24 for Ring,
- 24 for Schizonts, and
- 24 for Trophozoite.

Using image processing techniques, it was then performed feature extraction toward digital image data. Feature having to be extracted was values i.e. first and second order statistic feature. The image file was subsequently analyzed using an algorithm explained in the next section. All algorithms were evaluated using the same data. System flowchart proposed in this study could be seen in Figure 1.

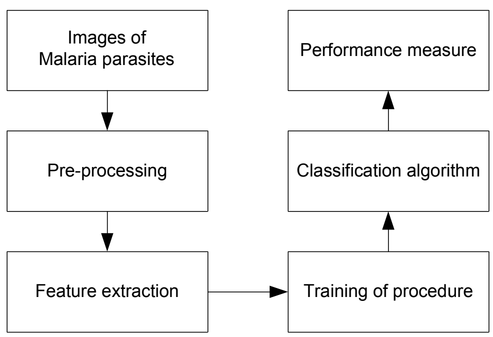

Figure 1. Block diagram of proposed system

2.1 Learning vector quantization (LVQ)

LVQ is one of methods to conduct guided classification (Kohonen, 1992; Bezdek and Pal, 1995; Karayiannis, 1997). LVQ consists of two layers, they are: 1) first layer is competitive one. This classifies input data by competition. 2) Second layer is linear one. This layer transforms subclasses resulted from first layer that was previously defined (target). Every subclass is represented by one neuron in competitive layer output and every class is represented by one neuron at linear layer output. Neuron in competitive layer output is usually called hidden neuron and neuron in linear layer output is called output neuron. Subclass in competitive layer is a result of this layer while the class in linear layer is a class defined by user Subclass (target).

The processing occurred in every neuron is finding out the closest distance between the input vector and the weight. During this term, weight vector (w) connects every neuron in input layer to hidden neuron on the output layer. Activation function on output layer will map weight vector from linear layer to the output layer.

<u>LVQ network training algorithm:</u>

1) Determine:
   a. Initial weight of j^{th} input variable toward i^{th} class that is W_{ij}, where i = 1, 2, ..., K and j = 1, 2, ..., m.
   b. Epoh maximum = Maxepoh.
   c. Learning rate parameter = α.
   d. Learning rate reduction = Decα.
   e. Allowable minimum learning rate = Minα.
2) Enter:
   a. Input data X_{ij} with i = 1, 2, ..., n and j = 1, 2, ..., m.
   b. Target as class, i.e. T_k with k = 1, 2, ..., n.
3) Set up epoh initial condition, that is epoh = 0.
4) Solve it if (epoh ≤ Maxepoh) and (α Minα)
   a. epoh = epoh + 1
   b. Solve it for i = 1 to n
      i. Determine J in such a way that $|X_i – W_j|$ minimum with j = 1, 2, , K.
      ii. Correct W_j with stipulation:
         - If T = C_j thus $W_j = W_j + \alpha (X_i – W_j)$
         - If T ≠ C_j thus $W_j = W_j – \alpha (X_i – W_j)$
   c. Subtracts α value (α value subtraction can be performed by $\alpha = \alpha$ - Decα or by $\alpha = \alpha$ * Decα.

Having been conducted training; it would be obtained final weights (W). It was these weights that would be used to conduct test and simulation with different data (different with the data used for training).

<u>LVQ network testing algorithm:</u>

1) Enter data that would be tested, for example X_{ij}
   with i = 1, 2, ..., np and j = 1, 2, ..., m.
2) Solve for i = 1 to np
   a. Determine J in such a way that $|X_i – W_j|$ minimum with j = 1, 2, ..., K.
   b. J is a class for X_i

2.2 Computational Complexity

An algorithm was not merely producing correct output, but it must be also effective. The correctness of an algorithm must be tested by certain input numbers to see the algorithm performance as time needed to run its algorithm and memory space required for its data structure.

Algorithm effectiveness was measured by several time number and memory spaces consumed to run the algorithm. Effective algorithm was the one minimizing time and space consumption. The implementation of an algorithm could be said properly if it met the formal criteria used to assess that algorithm i.e. algorithm effectiveness with its complexity. The magnitude used to explain algorithm time/space assessment model was by employing algorithm complexity.

There are two kinds of algorithm complexity, i.e. time complexity and space complexity. The time complexity of an algorithm is measuring total computation performed by computer when it completes a problem by applying the algorithm. The measurement referred to the number of calculation steps and time consumption of processing.

The time complexity is important to measure efficiency of algorithms. The time complexity of an algorithm is measured as a problem measurement function. The time complexity of an algorithm contains numeral expression and step amount needed as a function of problem measurement. Space complexity is associated with memory system required in program execution. On the Table 1, it's presented algorithm group based on its asymptotic time complexity.

Time and space consumption of algorithm depended on input dimension. Input dimension was symbolized as n. Having determined input dimension, the next step in measuring the time complexity would be figuring up the operation numbers already executed by algorithm thus it would be found its time complexity notation within n function, f(n).

To measure time consumption of an algorithm was by directly executing that algorithm in a computer, and then calculating the time duration consumed to complete a problem with various n. The algorithm computation result was then compared to its time complexity notation to find its algorithm efficiency.

Table 1. Time complexity functions

Algorithm group	Name
O(1)	Constant
O(log n)	Logarithmic
O(n)	Linear
O(n log n)	n log n
$O(n^2)$	Quadratic
$O(n^3)$	Cubic
$O(2^n)$	Exponential
O(n!)	Factorial

3. RESULT AND DISCUSSION

Malaria parasite pattern recognition was started by processing digital image of each class. Image processing began by minimizing noise factor as a result of image acquisition, extraction, and feature choosing.

The feature employed in this experiment was a combination of RGB image first and second order feature. However, there were only four best features resulted from finding feature by applying least square error that would be used in malaria parasite pattern recognition process by using the method already expressed in the earlier section. The best feature of the selection result would be mean feature of green, blue and red band as well as contrast feature from blue band.

Table 2, it would be found out that the final weight for every algorithm was different. Wright was the result of adaptation occurred within the data environment determined, in this case was feature of each class from several initial choices and used iterative procedure.

The final quality of the method actually determined the accuracy in recognizing the malaria parasite type itself. Figure 3 suggested the result of the malaria parasite recognizing system. The LVQ algorithm had a performance in recognizing the malaria parasite type with 92% accuracy level. The recognition successful level was influenced by the number of training data set. In addition to successful level testing in recognizing malaria parasite, pattern recognition system was also tested in terms of time consumption in its recognition process.

At the similar hardware specification, the most influencing things toward process time was the active number of software or program simultaneously operating at the testing time. Therefore, it's applied Big (O) theorem to examine system computation time constructed in real time condition. The testing was performed by giving test data input at the minimum number to maximum dataset number, that was 80 malaria image sample data set.

Figure 3 showed the process time comparison of the algorithms used. From the test results, it was obtained mean value of testing time for LVQ algorithm 2.03 second. The regression result of testing time distribution for this algorithm had equation 5.275e-5x+2.03, meaning that big (O) value of this algorithm was O (n). This means that process time was linear; it wasn't influenced by the number of datasets tested.

Table 2. Experiment result of the LVQ classification method

	Result		
	Final weight		Accuracy (%)
Gametocyte	194,2	2.077,0	95
Schizonts	221,5	509,2	92
Ring	201,1	1.445,1	87.5
Trophozoite	201,3	1.167,2	92
Average accuracy			91.625

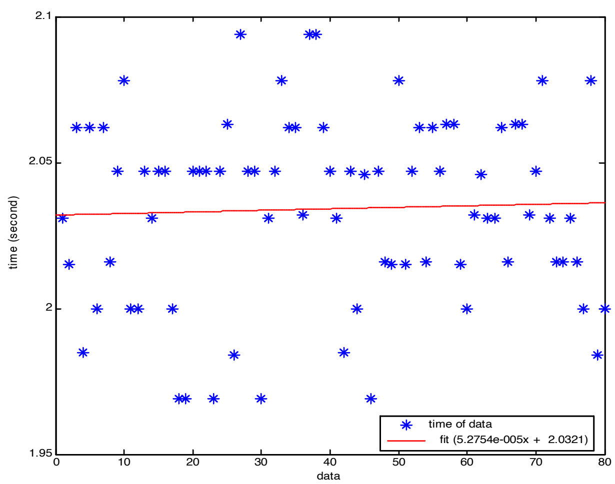

Figure 3. Real time test graph

4. CONCLUSION

From computation test result of all three aforementioned malaria parasite pattern recognition algorithms, it was found that the LVQ algorithm had better performance in terms of recognizing the malaria parasite type and it had faster computation time complexity. Therefore, this algorithm could be the primary candidate as the detecting method for malaria parasites.

5. REFERENCES

Arum, L. I., A. Purwanto, S. Arfi, H. Tetrawindu, M. Octora, and K. Surayah, 2006, Diagnostic Test of Plasmodium Malaria by Immunochromatographic Method Compared to Microscopic Examination: Indonesian Journal of Clinical Pathology and Medical Laboratory, v. 12, no. 3, p. 118–122.

Bezdek, J. C., and N. R. Pal, 1995, Two soft relatives of learning vector quantization: Neural Networks, v. 8, no. 5, p. 729–743.

Díaz, G., F. A. González, and E. Romero, 2009, A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images.: Journal of biomedical informatics, v. 42, no. 2, p. 296–307, doi:10.1016/j.jbi.2008.11.005.

J. Somsekar, 2011, Computer Vision for Malaria Parasite Classification in Erythrocytes: International Journal on Computer Science and Engineering (IJCSE), v. 3, no. 6, p. 2251–2256.

Karayiannis, N. B., 1997, Learning vector quantization: A review: International Journal of Smart Engineering System Design, v. 1, no. 1, p. 33–58.

Kekre, H. B., and T. K. Sarode, 2009, Centroid Based Fast Search Algorithm for Vector Quantization: p. 73–83.

Khan, M. I., B. Acharya, B. K. Singh, and J. Soni, 2011, Content Based Image Retrieval Approaches for Detection of Malarial Parasite in Blood Images: International Journal of Biometrics and Bioinformatics, v. 5, no. 2, p. 97–110.

Kohonen, T., 1992, New developments of learning vector quantization and the self-organizing map: Symposium on Neural Networks: Alliances and Perspectives in Senri, Osaka, Japan.

Le, M.-T., T. R. Bretschneider, C. Kuss, and P. R. Preiser, 2008, A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears.: BMC cell biology, v. 9, no. 1, p. 15, doi:10.1186/1471-2121-9-15.

Ma, C., P. Harrison, L. Wang, and R. L. Coppel, 2010, Automated estimation of parasitaemia of Plasmodium yoelii-infected mice by digital image analysis of Giemsa-stained thin blood smears.: Malaria journal, v. 9, no. 1, p. 348, doi:10.1186/1475-2875-9-348.

Makkapati, V. V, and R. M. Rao, 2011, Ontology-based malaria parasite stage and species identification from peripheral blood smear images.: Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, v. 2011, p. 6138–41, doi:10.1109/IEMBS.2011.6091516.

Mandal, S., A. Kumar, J. Chatterjee, M. Manjunatha, and A. K. Ray, 2010, Segmentation of blood smear images using normalized cuts for detection of malarial parasites, *in* 2010 Annual IEEE India Conference (INDICON): IEEE, p. 1–4, doi:10.1109/INDCON.2010.5712739.

Premaratne, S. P., N. D. Karunaweera, and S. Fernando, 2006, A Neural Network Architecture for Automated Recognition of Intracellular Malaria Parasites in Stained Blood Films: p. 4–7.

Sedgewick, R., and P. Flajolet, 2013, An Introduction to The Analysis of Algorithms: New Jersey, Addison-Wesley.

Sio, S. W. S., W. Sun, S. Kumar, W. Z. Bin, S. S. Tan, S. H. Ong, H. Kikuchi, Y. Oshima, and K. S. W. Tan, 2007, MalariaCount: An image analysis-based program for the accurate determination of parasitemia: Journal of Microbiological Methods, v. 68, no. 1, p. 11–18.

WHO, 2014, Malaria Fact Sheet No.94: p. 3–9.