A Combined Genetic Algorithm and Conway's Game of Life for Printed Lanna Character Recognition

Krisda Khankasikam, *Senior Member, IACSIT*

Abstract—**This research is a study of using genetic algorithms through Conway's game of life to solve the problem on Lanna character recognition. Pattern recognition can be defined as the objects identification on the basis of information available about it. The problem on character recognition has been discussed to find out the best solution and various recognition methods have been implemented. One of the proposed methods is genetic algorithm which used in many application areas such as image processing, pattern recognition and machine learning. Genetic algorithms are powerful search techniques that applied natural selection and genetic operators to find the best result. Based on the survey of related research, this research is provided to investigate the possibility of using genetic algorithms on Lanna character recognition. The proposed framework consists of three modules: data preparation, training by using genetic algorithms and recognition. The experiments are carried on 1,200 Lanna characters, which randomly selected from documents written in Lanna language. The system performs the number of match 1,142 characters or 95.15%, while mismatch 27 characters or 2.25% and rejecting 31 characters or 2.58%.**

Index Terms—**Genetic algorithms, conway's game of life, lanna character, pattern recognition.**

I. INTRODUCTION

Pattern recognition is concerned with an attribute of human being as well as other living organisms. Human can recognize the voice of known person, read handwriting, analyze fingerprints and distinguish a countenance. A human being is a very sophisticated information system because human possesses a superior pattern recognition capability [1]. Many definitions of pattern recognition are proposed, pattern recognition can be defined as a process which leads to a decision. Moreover, pattern recognition can be defined as an area of science concerned with identifying objects on the basis of information available about it. The character recognition problem has been approached in many ways and various recognition methods have been suggested. Some methods are developed for the character recognition but most of them are borrowed from other fields of pattern recognition such as signal and image processing. The printed character and handwritten recognition are the first practical application in the field of pattern recognition [2].

Genetic algorithms are algorithm that can be easily applied to solve difficult problem because of its generic structure. It

is used in many application areas such as image processing, pattern recognition and machine learning [3], [4]. Genetic algorithms are powerful search techniques that applied natural selection and genetic operators to find the best result. Its power comes from its ability to combine good pieces from different solutions and assemble them into a single super solution [5], [6]. Genetic algorithms are generated initial population of solution called individuals or chromosomes then the solutions are evaluated by using fitness function. The algorithm creates new generations of population by genetic operations such as reproduction, crossover and mutation. The next generation consists of the possible survivors and the new individuals obtained from the previous population by the genetic operations. The result of using genetic algorithms for character recognition in other language yields accuracy rate at 75-90%.

The research on character recognition in Thailand has been active for a decade [7]. Various recognition techniques are operated such as recognition based on detecting an important structure of character [8]-[11] recognition based on fuzzy rough sets [12], [13] recognition based on linear interpolation approach [14] and recognition based on neural network [15]. These methods have its own advantage and disadvantage, and can not be identified to be the best. A proper approach for the entire solution of the recognition problem has not been found and most of them are focused on Thai character recognition. Based on the survey of related research, the author is provided to conduct the research in Lanna character [16] recognition which applied genetic algorithm to solve problem.

The rest of this paper is organized as follow. Section II provides a brief related works. Section III describes proposed methodology, followed by a description of the experiment in Section IV. Conclusion is drawn in Section V.

II. BACKGROUND MATERIALS

This section gives a brief description of related works consisted of Lanna character, Genetic Algorithms and Conway's rules of life.

A. Lanna Language

This section gives a brief overview of Lanna language. The Lanna character is used to write Northern Thai language. The Lanna character set consists of 49 consonants, 13 vowels and 18 mixed vowel. Almost Lanna character composes of loops which are attached to straight lines and circular arcs. Almost Lanna characters are written by using stroke. Fig. 1 shows the example of Lannna language.

Based on Fig. 2, The Lanna sentence structure is divided into 3 main levels as in Latin family language. Additionally,

Lanna language has 2 sublevels in the upper level. The big difference between Lanna and Latin is Lanna does not have space between words. Moreover, Lanna characters have more complex structure than Latin.
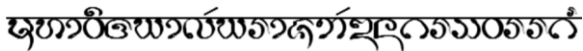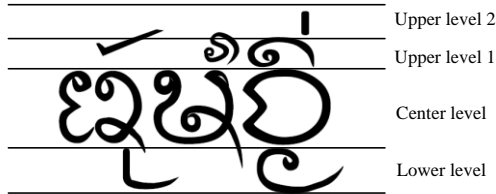

Fig. 1. The example of Lanna language.

Fig. 2. Lanna sentence structure.

In this research, 49 Lanna characters with 8 fonts and different 10 character sizes are selected to create learning database in data preparation and training by using genetic algorithms module.

B. Genetic Algorithms

In this section, an overview of genetic algorithms is explained. Next, the different operators of genetic algorithms are explored and finally the general outline of genetic algorithms is shown by using a flowchart.

Genetic Algorithms (GA) are a probabilistic search which is first suggested by John Holland in 1975 [17]. The general genetic algorithms work with three operators: selection, crossover and mutation [18]. The functioning of these three operators is shown in Fig. 3. These operators work on the data sets of the algorithm, which called chromosomes. Another important part of the algorithm is the fitness function. This function calculates the fitness of a chromosome using the genes of the chromosome. The goal of a genetic algorithm is to optimize this fitness function and find the individual that has the "best" fitness value.
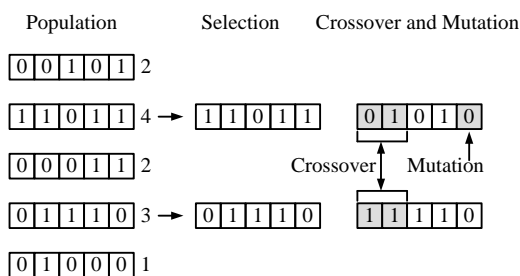

Fig. 3. The genetic algorithms.

Fig. 3 shows an example of a genetic algorithm. The process starts with a population of 5 chromosomes. The genes of the chromosomes are binary encoded which contained the only values 0 and 1. The values behind the chromosomes correspond to their fitness. In this example the fitness-function just counts the number of ones in the chromosome. After calculating the fitness of all the chromosomes in the population pool, two chromosomes must be selected for reproduction. This selection process is often random but with a bias for the chromosomes with higher fitness. Following the selection process the crossover and mutation operators will work on the two chromosomes. The

crossover operation exchanges a number of genes from one chromosome with another. The mutation operator just changes the value of one gene in a chromosome. In the right part of Fig. 3 the left two genes of the chromosomes are exchanged and the mutation occurs on the right side of the upper chromosome. Crossover and mutation do not always happen. The general outline of the genetic algorithm is given in Fig. 4.
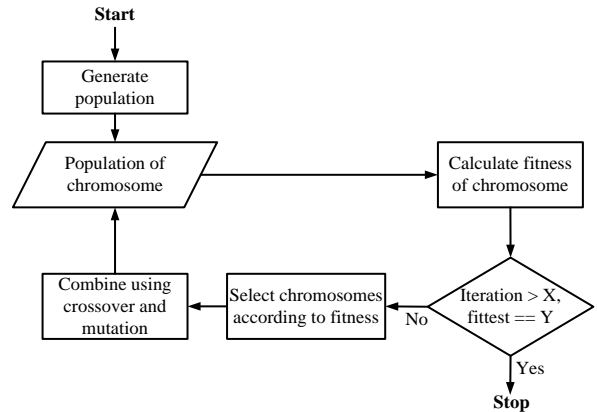

Fig. 4. Flowchart of the genetic algorithms.

In this research, mutation operator, one of three main operators of genetic algorithms, is applied to combine with Conway's rule of life. The details of mutation rate and characteristic for solving Lanna character recognition will be explained in the next section.

C. Conway's Game of Life

Conway's game of life was developed by John Horton Conway in 60s [19]. It is a set of rules which defined to control mutation behavior of two dimension cell in checkerboard. The basic idea is to start a simple configuration of cell then observe change by applying Conway's genetic law for survivals, births and deaths [20], [21]. The rules are:

Rule 1 Survival: if a live cell has two or three live neighbors, it survives.

Rule 2 Death: if a live cell has less than two or more than three live neighbors, it dies.

Rule 3 Birth: if a dead cell has exactly three live neighbors, it is born.

These rules create the pattern of cells that will be survival or die or birth, depending on a description of eight neighbor cells, N1 to N8, as shown in Fig. 5.
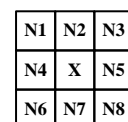

Fig. 5. The neighbor cells position used for consideration.

In case that cell X has a status survival or birth, based on Conway's rules, an effect on cell X may be the following:

Effect 1: if there are 2 or 3 cells around cell X has status survival or birth, cell X will be survived.

Effect 2: if there are fewer than two cells or more than three cells round cell X has status survival or birth, cell X will be died.

In case that cell X has a status death, its status will be

changed to birth when there are three cells around cell X have status survival or birth.
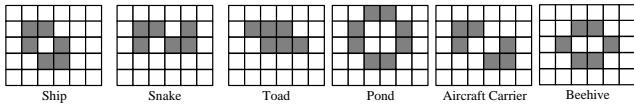

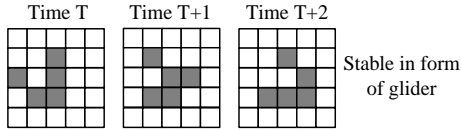Fig. 6. The examples of stable state of Conway's game of life.

Fig. 7. The examples of Conway's game of life.

The stable state in Conway's game of life is a state that contained only cell patterns which will not change any more. The some examples of cell pattern in stable state are illustrated in Fig. 6 and Fig. 7 shows the examples of Conway's game of life.

Conway's game of life has a capability to generate a new identity pattern, which is a stable state, from predefined pattern. In this research, the idea of Conway's rule of life is applied by assuming that the each line of Lanna character, in form of 30 X 30 pixels matrix, is a chromosome. In initial state, black pixels are considered as survival and birth status and white pixels are considered as death status.

III. METHODOLOGY

This section describes the proposed methodology of the Lanna character recognition which composed of three modules: data preparation, training by using genetic algorithms and recognition. The overview of proposed methodology is shown in Fig. 8.
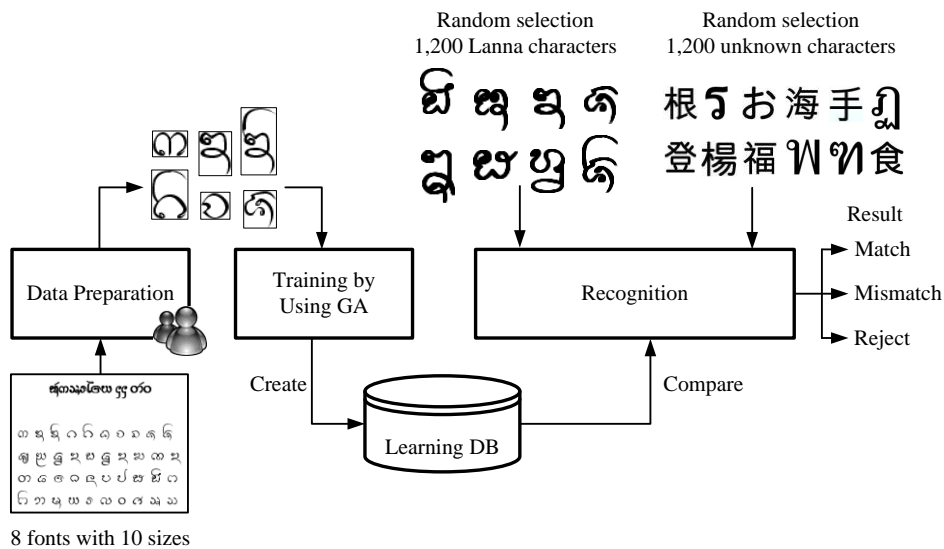

Fig. 8. The system overview.

A. Data Preparation

This module proposes process to capture Lanna character from documents written by Lanna language. The input data, inform of scanned image, must have black characters on the white background. Each character is manually segmented through the Microsoft Paint as shown in Fig. 9. The 49 Lanna character derived from 8 fonts type with 10 different sizes are selected to be input data in data preparation process. The outputs of this module, which are the inputs of next module, are individual font in form of image.

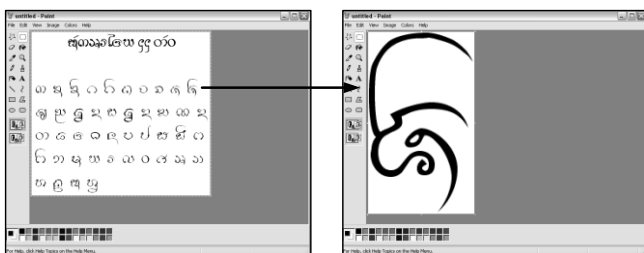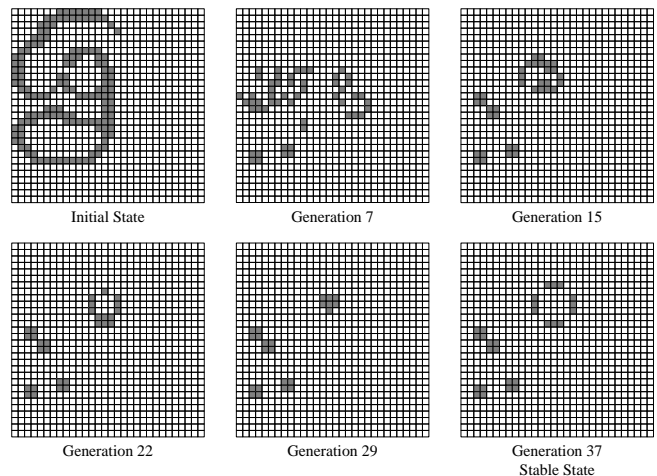
Fig. 9. Data preparation through Microsoft Paint.

Fig. 10. Lanna character mutation through Conway's game of life.

B. Recognition

This module attempts to recognize the unknown character through software which implemented for this research. The software will be compared the stable state of unknown

character with all stable state in the database. The result of this module can be match, mismatch and reject.

IV. EXPERIMENT

The experiments on Lanna character recognition by using genetic algorithms are performed on 8 fonts type composed of lannawor, Cr-Doitong, Cr-Father, Cr-Fak, Cr-Insom, Cr-Udom, Cr-Maekorn and Cr-Kern. The 1,200 characters from 10 different font sizes, from 8 to 26, are randomly selected to test the system performance. The system can recognized (match) 1,142 characters or 95.15%, while mismatching 27 character or 2.25% and rejecting 31 characters or 2.58% as shown in Table I.

Moreover, the experiments are performed on 1,200 unknown characters with 12 different font types which derived from Chinese, Japanese and Thai characters. Regarding to select unknown character, Ming LiU, MS Hei, MS Song and Sun ExtA are example of Chinese characters, while Hana Min Regular, Kochi Mincho, MS Mincho and MS UI Gothic are selected to example of Japanese characters, and AngsanaUPC, CordiaUPC, EurosiaUPC and Tahoma are example of Thai characters. The system informs the number of match 6 characters or 0.5%, while mismatching 11 characters or 0.92% and rejecting 1,183 characters or 98.58% as shown in Table II.

TABLE I: EXPERIMENTAL RESULT FOR LANNA CHARACTER.

Font Name	No. of Character	Results		
		Match	Mismatch	Reject
Lannawor	150	148	2	0
Cr-Doitong	150	142	6	2
Cr-Father	150	145	4	1
Cr-Fak	150	145	2	3
Cr-Insom	150	136	3	11
Cr-Udom	150	140	5	5
Cr-Maekorn	150	141	1	8
Cr-Kern	150	145	4	1
Total	**1,200**	**1,142**	**27**	**31**
Percentage	**100%**	**95.15%**	**2.25%**	**2.58%**

TABLE II: EXPERIMENTAL RESULT FOR UNKNOWN CHARACTER.

Font Name	No. of Character	Results		
		Match	Mismatch	Reject
Ming LiU	100	0	1	99
MS Hei	100	0	0	100
MS Song	100	0	0	100
Sun ExtA	100	0	1	99
Hana Min	100	0	0	100
Kochi Mincho	100	1	0	99
MS Mincho	100	0	1	99
MS UI Gothic	100	0	1	99
AngsanaUPC	100	1	2	97
CordiaUPC	100	1	1	98
EurosiaUPC	100	1	2	97
Tahoma	100	2	2	96
Total	**1,200**	**6**	**11**	**1,183**
Percentage	**100%**	**0.5%**	**0.92%**	**98.58%**

V. CONCLUSION

This paper has described the Lanna character recognition by using genetic algorithm which applied Conway's game of life as a mutation pattern. The Conway's game of life is applied to determine the status of cell that is survival, death

and birth within a 30 X 30 pixel matrix. The experiments are performed on 1,200 printed Lanna alphabets which adjusted sharpness of character line. The system performs the level of match, mismatch and reject at 95.15%, 2.25% and 2.58%, respectively.

For the future work, the system will focus on applying other algorithms besides genetic ones, for example, case-base reasoning and neural network. In addition, there should be a variety and number of character sample patterns to increase the possibility of recognition.

REFERENCES

[1] C. A. Perez, C. A. Salinas, P. A. Estevez, and P. M. Valenzuela, "Genetic design of biologically inspired receptive fields for neural pattern recognition," *IEEE Tran on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 33, issue 2, pp. 258-270, April 2003.
[2] M. A. Abed, A. N. Ismail, and Z. M. Hazi, "Pattern recognition using genetic algorithm," *Int. Jour. of Computer and Electrical Engineering,* vol. 2, no. 3, pp. 583-588, June 2010.
[3] C. Pornpanomchai and M. Daveloh, "Printed thai character recognition by genetic algorithm," in *Proc. Int. Conf. Machine Learning and Cybernetics,* Hong Kong, pp. 3354-3359, 2007.
[4] X. Wei, S. Ma, and Y. Jin, "Segmentation of connected Chinese characters based on genetic algorithm," in *Proc. 8th Int. Conf. Document Analysis and Recognition,* pp. 645-649, 2005.
[5] B. Bontempi and A. Marcelli, "A genetic learning system for on-line character recognition," in *Proc. 12th IAPR Int. Conf. Pattern Recognition,* Jerusalem, pp. 83-87, 1994.
[6] B. Bontempi and A. Marcelli, "Machine learning and genetic algorithms: an application to character recognition," in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics,* San Antonio, pp. 2225-2230, 1994.
[7] H. T. Koanantakool, T. Karoonboonyanan, and C. Wutiwiwatchai, "Computers and the Thai language," *IEEE Annals of the History of Computing,* vol. 31, issue 1, pp. 46-61, 2009.
[8] A. Jirayusakul, "Apply the combination of multiple classifiers with the SGNG algorithm for Thai printed character recognition," in *Proc. 8th Int. Symp. Natural Language Processing,* Bangkok, 2009, pp. 79-82.
[9] B. Kruatrachue and P. Piyatrakul, "Automatic Thai and English fonts identification without character recognition," in *Proc. IEEE Pacific Rim Conf. Communications, Computers and Signal Processing,* Victoria, pp. 603-606, 2001.
[10] S. Tangwongsan and C. Sumetphong, "Optical character recognition techniques for restoration of Thai historical documents," in *Proc. Int. Conf. Computer and Electrical Engineering,* Phuket, pp. 531-535, 2008.
[11] S. Tangwongsan and O. Jungthanawong, "A refinement of stroke structure for printed Thai character recognition," in *Proc. 9th Int. Conf. Signal Processing,* Beijing, pp. 1504-1507, 2008.
[12] S. Mitatha, K. Dejharn, F. Chevasuvit, B. Chankuang, and W. Kasemsiri, "Experimental results of using rough sets for printed Thai characters recognition," in *Proc. IEEE Region 10 Int. Conf. Electrical and Electronic Technology,* pp. 331-334, 2001.
[13] W. Kasemsiri and C. Kimpan, "Printed Thai character recognition using fuzzy-rough sets," in *Proc. IEEE Region 10 Int. Conf. Electrical and Electronic Technology,* pp. 326-330, 2001.
[14] P. Jamjuntr and Dejdumrong, "Thai font type recognition using linear interpolation analysis," in *Proc. 6th Int. Conf. Computer Graphics, Imaging and Visualization,* Tianjin, pp. 406-409, 2009.
[15] B. Kijsirikul, S. Sinthupinyo, and A. Supanwansa, "Thai printed character recognition by combining inductive logic programming with backpropagation neural network," in *Proc. IEEE Asia-Pacific Conf. Circuits and Systems,* Chiang Mai, pp. 539-542, 1998.
[16] K. Khankasikam, "Evaluated conway's game of life in lanna character detection," in *Proc. Int. Conf. Information and Multimedia Technology,* Hong Kong, pp. 1-4, 2010.

[17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Prentice Hall, ch. 4, 2003.

[18] Y. Kimura, A. Suzuki, and K. Odaka, "Feature selection for character recognition using genetic algorithm," in *Proc. 4th Int. Conf. Innovative Computing, Information and Control*, Kaohsiung, 2009, pp. 401-404.

[19] M. Gardner, "Mathematical games: The fantastic combinations of John Conway's new solitaire game "life"," *Scientific American*, vol. 223, pp. 120-123, October 1970.

[20] M. R. Lauer, P. A. Mitchem, and R. A. Gagliano, "Resource optimization and self-interest: variations on the game of life," in *Proc. 28th Annu. Simulation Symp.*, Phoenix, pp. 136-143, 1995.

[21] T. Fawcett, "Data mining with cellular automata," *ACM SIGKDD Explorations Newsletter*, vol. 10, issue 1, pp. 32-39, June 2008.

Krisda Khankasikam was born in Thailand, on September 8, 1980. He received the B.Eng. degree in computer engineering from Naresuan University, Phitsanulok, Thailand, in 2002, the M.Eng. degree in computer engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 2005 and the Ph.D. degree in knowledge management from Chiang Mai University, Chiang Mai, Thailand, in 2010.

He is an Assistant Professor of Field of Computer and Information Technology, Department of Applied Science, Faculty of Science and Technology, Nakhon Sawan Rajabhat University, Thailand. His research interests are pattern recognition, information retrieval and knowledge management. Dr. Krisda is a senior member of International Association of Computer Science and Information Technology (IACSIT), Singapore.