International Journal of Scientific & Engineering Research Volume 9, Issue 12, December-2018
ISSN 2229-5518

OPTIMIZATION SUPPORT VECTOR MACHINE PARAMETERS WITH GENETIC ALGORITHM ON TEXT INDEPENDENT VOICE RECOGNITION

Ikra Dewantara, Agus Buono, Bagus Sartono

Abstract— Voice recognition using computer is made possible do the unique characteristics and frequencies that each voice possesses. Voice recognition based independent text is voice recognition using free words. The purpose of this research are to modeled voice recognition using Support Vector Machine (SVM) and optimize SVM kernel using Genetic Algorithm (GA). The data are from 5 people that each people record 985 words. We use Mel Frequency Cepstrum Coefficient (MFCC) as feature extraction and Vector Quantization (VQ) to compress data as input in SVM. We tested with 2 methods using determined parameters kernel value and optimized parameters kernel with GA. The best identification using determined parameters kernel is 98.40% with RBF kernel. The best identification using optimized parameters kernel with GA is 99.20% with RBF kernel.

Index Terms— Genetic algorithm, Mel Frequency Cepstrum Coefficient (MFCC), Support Vector Machine (SVM), Vector quantization, Voice identification, Voice recognition.

1 INTRODUCTION

Biometric recognition is one of human recognition method done by machine. This recognition is done for authentication process in the system. Authentication system that were developed in this day among them are iris scanner, face recognition, fingerprint scanner, and voice recognition.

According Kinnunen (2010), voice recognition can be applied into authentication system because physically every speaker has different throat shape, larynx, and other organs to produce a voice. The difference from habit side such as accent, rhythm, intonation, and pronouncation can be another feature from every speaker.

Voice recognition can be divided into 2 categories, i.e text dependent and text independent. Text dependent voice recognition attached to word that have been previously defined and used in training process and verification. While text independent voice recognition isn't attached to certain text. The system has to recognize any text when verification. (Rydin 2001)

The first step in voice recognition is to change the voice in analog form into digital form. The result is a vector representation in bulk without eliminating the feature from that voice. Therefore, it's needed some form of a feature extraction technique to change voice vector into feature vector without reducing the voice characteristic. According to Narang (2015), mel-frequency cepstrum coefficients (MFCC) can represent feature extraction signal much better than linear prediction ceptrum coefficient (LPCC), relativespectral, and probabilistic linear discriminate analysis (PLDA).

The feature extraction result in text dependent and text independent voice recognition have different data size. Text independent voice recognition has a larger data size because speakers are free to say any word as long as they want. Larger feature extraction data will make modeling process take a longer time. According to Rydin (2001) to reduce the data size there are 3 ways that can be used such as long-term statistics-based systems, vector quantization methods, and ergodic HMM-based methods. Smith (2012) too explain that reducing data size using long term statistic-based systems can eliminate a lot of important information because the average value from the data was taken. According to Kekre (2008) vector quantization has potential to reduce data while keeping the quality of the data.

Voice recognition research has been done by Smith (2012) using SVM as sound pattern recognition. According to Smith (2012) SVM has been used so many times and produce a fine sound pattern recognition. SVM can solve grouping 2 class problem perfectly using 2 linear areas division method. As for the obstacles faced by Smith's research are pattern recognition using SVM in solving grouping non-linear class problem and the modeling time that took a long time when facing a larger amount of data. Large data processing constraints too faced by Salomon (2001). According to Salomon (2001) the key to increase SVM performance by solving large data problem because that data quite consuming the memory.

According to Nijhawan (2014) in his research entitled Speaker Recognition Using MFCC and Vector Quantization, MFCC result data with big capacity can be reduced with vector quantization while keeping the quality of the data. This small sized vector is useful as input data for SVM in order to shorten the modeling time while keeping the level of accuracy.

Voice recognition research using SVM too has been done by

- *Ikra Dewantara is currently pursuing masters degree program in computer science in Bogor Agricultural University, Indonesia, PH+628111189895. E-mail: dewantaraikra@gmail.com*
- *Agus Buono is currently lecturer in department of computer science in Bogor Agricultural University, Bogor, Indonesia.*
- *Bagus Sartono is currently lecturer in department of statistics in Bogor Agricultural University, Bogor, Indonesia.*

Mezghani (2010) entitled Evaluation of SVM Kernels and Conventional Machine Learning Algorithms for Speaker Identification. According to Mezghani (2010) one of the SVM problems is how to choose the right kernel in accordance with the dataset. Each kernels have their own parameterss that can be adjusted with the dataset to increase the accuracy. According to Huang (2006) Genetic Algorithm (GA) can increase accuracy of a algorithm. Huang's research entitled A GA-based feature selection and parameterss optimization for support vector machine discuss about using genetic algorithm as optimazation on SVM algorithm in processing voice identification.
Based on the introduction, then this study aims to identifiy voice based on text independent using MFCC feature extraction, the algorithm on vector quantization preprocessing stage, classification using SVM, and optimize SVM parameterss using GA. Expected result from this study are to get kernel and a good parameters value to produce the highest accuracy.

2 Reseach Method

This study consists of several stages, namely data collection, preprocessing, feature extraction, distribute training and test data, voice modeling, testing, and SVM kernel parameters optimization using GA evaluation. The research flow presented on Figure 1 and Figure 2.

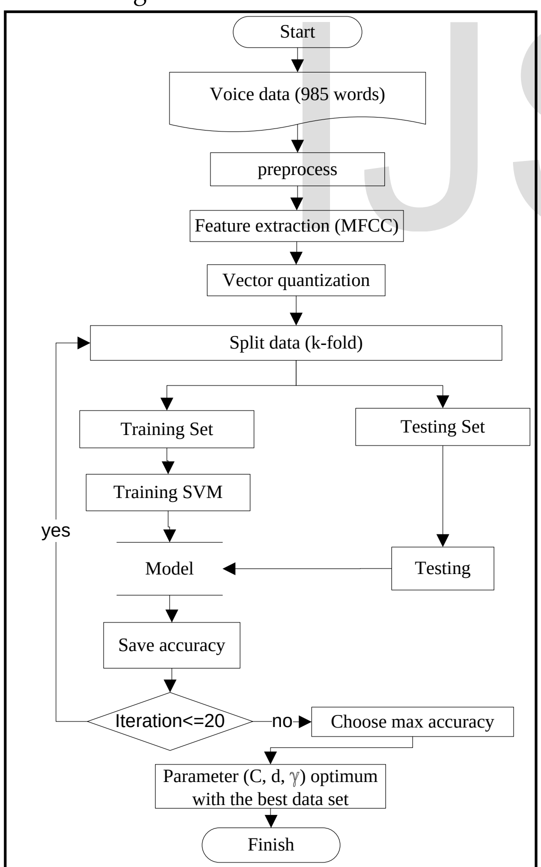

Fig. 1. Reseach framework voice recognition used SVM with fixed parameters.

2.1 Data Collection

The data that used in this research are voice recording data from 5 speakers (3 women and 2 man) with an age range from 25 until 35 years old. That data recorded in Computational Intelligence Laboratory Computer Science FMIPA IPB Bogor using headphone that connected to the computer. The application used to record is Audacity 2.1.2. Each voice data consist 985 words in Indonesian language.

2.2 Data Preprocessing

Preprocessing stage consists of normalization and silent removal. Normalization stage is changing the data on range -1 until 1. This is being done to reduce existing amplitude differences when there is high pitch and low pitch volume differences.
Silent removal stage is a stage to erase silent voice. Silent voice is a voice condition without data (the volume at 0). The purpose of silent removal application so that the processed data contains full data (not a blank data).

2.3 Feature Extraction

The voice that has been normalized and the silent voice has been removed, then the features extracted using MFCC. As for MFCC parameters that used in this study among them are: time frame 40, sampling rate 11000 Hz, overlap 0.5, and the number of coefficients that used on every frame are 13.

2.4 Vector Quantization (VQ)

Vector quantization (VQ) technique used to reduce data size that were big enough while keeping the quality of the data. Each cluster has center point (centroid). Centroid on VQ called codeword. Collection of codeword called codebook. The output from using VQ technique in the form of vector with the size (13 x 100).

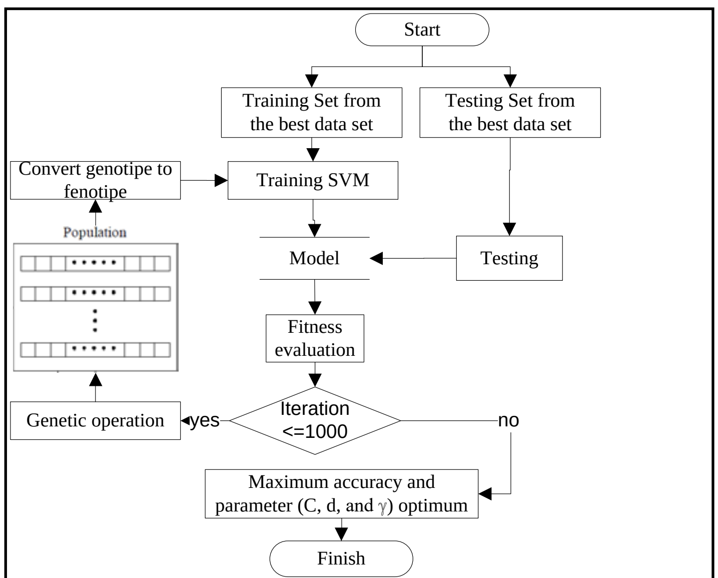

Fig. 2. Reseach framework voice recognition used optimization parameters support vector machine with genetic algorithm.

2.5 Data Distribution

Training data and test data distributed using k-fold cross validation. This study uses k=4 value. Proportion for training and test data is 75%:25%. Data on fold that produce the highest accuracy will be used on SVM parameters optimization using GA.

2.6 Feature Extraction

SVM modeling is being done using one vs all method. Every class will be paired with their class, while the other class considered isn't part of the class. The first modeling is being done to test SVM using determined parameters. The test was repeated 20 times to search the best data. The second modeling is optimizing SVM parameters using the data that were obtained previously on the first modeling. Optimizing SVM parameters testing too were repeated 10 times to find the highest accuracy. Some commonly used kernels include:

1. Linear kernel

$$K(u,v)=(u.v) \tag{1}$$

2. Polynomial kernel

$$K(u,v)=(u.v+1)^d \tag{2}$$

3. Radial basis function (RBF) kernel

$$K(u,v)= \exp(-\gamma|u-v|^2) \text{ , } \gamma>0 \tag{3}$$

with:
u = training data
v = class on training data
d and γ are kernel parameters

2.7 SVM Optimization using GA

SVM optimization is being done with the purpose to find the highest accuracy value from various kinds of SVM kernel parameters combination that were tested. Some kernels that tested among them are kernel linear, kernel polynomial, and kernel RBF with parameters such as C, γ, and d.

2.8 Evaluation

Evaluation stage is being done to calculate total accuracy average from each model. The result from the accuracy will be compared between SVM modeling accuracy and SVM with optimization using GA accuracy. Overall, accuracy were calculated based on equation 4 and 5:

$$\text{Accuracy determined parameters} = \frac{\sum \text{test voice correctly identified}}{\sum \text{voice in actual class}} \times 100\% \tag{4}$$

$$\text{Accuracy optimization SVM using GA} = \frac{\sum \text{test voice correctly identified}}{\sum \text{voice in actual class}} \times 100\% \tag{5}$$

3 Results and Discussion

3.1 Data Preprocessing

Data preprocessing begins with eliminating blank voice then ended with data normalization. The data comparison before and after preprocessing from side time presented on Table 1.

TABLE 1
DATA COMPARISON BEFORE AND AFTER PREPROCESSING

Speaker	Duration before preprocessing (minute)	Duration after preprocessing (minute)
1	07:55	05:30
2	08:21	05:39
3	07:09	05:27
4	06:21	05:18
5	08:38	06:20

3.2 Feature Extraction

Feature extraction process will take certain values as characteristic from each speakers. Feature extraction in this study uses MFCC. Parameters that were used among them are time frame 40, sampling rate 11000Hz, overlap 0,5 and the number of coefficients that used on every frame are 13. The results of MFCC application as feature extraction presented on Table 2.

TABLE 2
COMPARISON OF NORMALIZED AND MFCC DATA

Speaker	Normalization result (row x column)	MFCC feature extraction result (row x column)
1	16.769.486 x 1	13 x 76224
2	14.032.424 x 1	13 x 63783
3	14.959.683 x 1	13 x 67998
4	14.431.084 x 1	13 x 65595
5	14.569.394 x 1	13 x 66224

3.3 Vector Quantization (VQ)

Based on Table 2, the data dimension for the 5 data are very large. Large data dimensions will slow down the training process as presented on Table 3.

Table 3
COMPARISON TRAINING TIME

	Amount of Data	Linear	Polynomial	RBF
Time	100	1.14 sec	0.71 sec	0.68 sec
	1000	28.81 sec	12.92 sec	6.81 sec

Modeling on Table 3 is 5 parameters value modeling. On determined parameters testing, linear kernel will model 5 parameters with 4-fold 20 trial (total 400 modeling). Each polynomial kernels and RBF will model 25 parameters with 4-fold 20 trial (2000 modeling). Therefore, the data that used in this study is 13x100. MFCC and VQ results data comparison presented on Table 4.

Table 4
COMPARISON OF MFCC DATA AND QUANTIZED DATA

Speaker	MFCC feature extraction result (row x column)	VQ result data (row x column)
1	13 x 76224	13 x 100
2	13 x 63783	13 x 100
3	13 x 67998	13 x 100
4	13 x 65595	13 x 100
5	13 x 66224	13 x 100

ISSN 2229-5518

3.4 Data Grouping

Based on Table 4 on VQ result data column, if we sum each speaker data we will have total data 13 x 500. The coefficient value indicated by the value 13, while total data indicated by the value 500. Training data and test data distribution are using a ratio of 75% for training data and 25% for test data, so that there are 375 training data and 125 test data obtained.

3.5 SVM Testing using Determined Parameters

Before optimizing SVM parameters, the first test is being done to search the best accuracy using determined SVM parameters. Parameters value that used for each parameters presented on Table 5.

TABLE 5
DETERMINED PARAMETERS VALUES

Kernel	C	d	γ
Linear	10		
	20		
	30	-	-
	40		
	50		
Polynomial	10	2	
	20	4	
	30	6	-
	40	8	
	50	10	
RBF	10		10
	20		20
	30	-	30
	40		40
	50		50

Based on Table 5 above, each kernels will be tested using various kinds of parameters combination. The results of accuracy testing on three kernels above using determined parameters value presented on Table 6.

TABLE 6
RESULT OF VOICE RECOGNITION USING DETERMINED PARAMETERS

	Linear	Fold	Polynomial	Fold	RBF	Fold
Trial 1	86.40%	F4	92.80%	F3	95.20%	F4
Trial 2	92.00%	F4	92.80%	F3	94.40%	F4
Trial 3	84.80%	F3	91.20%	F4	96.80%	F3
Trial 4	94.40%	F3	91.20%	F2	96.00%	F2
Trial 5	88.00%	F2	95.20%	F2	96.80%	F1
Trial 6	92.80%	F2	96.80%	F2	96.00%	F2
Trial 7	90.40%	F4	89.60%	F4	96.00%	F4
Trial 8	89.60%	F1	90.40%	F3	95.20%	F3
Trial 9	88.00%	F1	91.20%	F1	97.60%	F1
Trial 10	87.20%	F4	93.60%	F4	**98.40%**	**F2**
Trial 11	91.20%	F2	90.40%	F3	96.00%	F1
Trial 12	94.40%	F1	90.40%	F4	95.20%	F2
Trial 13	87.20%	F2	91.20%	F4	96.00%	F2
Trial 14	90.40%	F4	92.00%	F4	94.40%	F3
Trial 15	94.40%	F4	91.20%	F3	96.00%	F2
Trial 16	91.20%	F3	92.80%	F3	95.20%	F2
Trial 17	87.20%	F4	88.80%	F4	96.80%	F1
Trial 18	89.60%	F4	88.80%	F4	94.40%	F1
Trial 19	88.00%	F1	92.00%	F1	95.20%	F3
Trial 20	92.80%	F4	92.80%	F1	95.20%	F4

Based on Table 6, of the 20 repeat experiments the linear kernel highest accuracy was 94.40% on 4th trial 3rd fold, 12th trial 1st fold, and 15th trial 4th fold. Polynomial kernels highest accuracy was 96.80% on 6th trial 2nd fold. RBF kernels highest accuracy was 98.40% on 10th trial 2nd fold. Based on the three kernels comparison, then the data that used for optimize SVM parameters using genetic algorithm is the data on 10th trial 2nd fold that produce the highest accuracy by 98.40%.

3.6 SVM Parameters Optimization using Genetic Algorithm

SVM optimization is being done with a purpose to search the highest accuracy value from various kinds of SVM kernel parameters combination that were tested. Some kernels that tested among them are linear kernels, polynomial kernels, and RBF kernels with parameters such as C, γ, and d. Fitness function in this study is accuracy. Chromosome shape on this study presented on Figure 3.

C	d	γ

Fig. 3. Chromosome design of SVM parameters optimization.

The second testing is optimizing SVM parameters is being done because the accuracy value using determined parameters felt haven't reached the maximum value. The first step SVM parameters optimization using GA begins with population initialization which consist of 3 individuals namely linear, polynomial, and RBF. The number of generations is limited to only 1000 generations.
Genes on each individual is a conversion from binary number to decimal number or in GA known as genotype – phenotype conversion. When a population is raised, the number that used is binary number. When testing fitness value the number that used is decimal number, therefore it's needed to convert from binary number into decimal number in determined range. According to Huang (2006) genotype to phenotype conversion can be done using equitation (6)

$$\square = \square\square\square_{\square} + \frac{\square\square\square_{\square} - \square\square\square_{\square}}{2^{\square}-1}\ \square\ \square \qquad (6)$$

p = Phenotipe of bit string
minp = minimum value of the parameters
maxp = maximum value of the parameters
d = decimal value of bit string
l = length of bit string

After all values on each individual have been converted into phenotype form, the next step is to calculate fitness value on each individual. Deciding crossover probability by 0.9 is chosen in hope from 90% crossbreed individuals on each generations can produce a more superior fitness from the previous (parent) generation. Mutated probability by 0.1 is chosen in hope from 10% mutated individuals on each generations aren't damaging the best individuals. The results of SVM parameters optimization is presented on Table 7

Table 7
RESULT OF VOICE RECOGNITION USING DETERMINED PARAMETERS

	Linear			Polynomial			RBF		
	min	average	max	Min	average	max	min	average	Max
Trial 1	44.00%	75.57%	98.40%	39.20%	71.36%	91.20%	0.00%	89.47%	99.20%
Trial 2	45.60%	77.89%	98.40%	39.20%	70.34%	92.00%	0.00%	89.95%	99.20%
Trial 3	43.20%	77.02%	98.40%	39.20%	72.35%	91.20%	0.00%	89.02%	99.20%
Trial 4	43.20%	77.55%	97.60%	39.20%	68.39%	91.20%	0.00%	89.50%	99.20%
Trial 5	47.20%	78.95%	97.60%	39.20%	71.66%	91.20%	0.00%	90.85%	99.20%
Trial 6	44.80%	77.23%	97.60%	39.20%	69.51%	91.20%	0.00%	90.69%	99.20%
Trial 7	44.80%	78.98%	98.40%	39.20%	68.84%	92.00%	0.00%	90.59%	99.20%
Trial 8	44.80%	78.32%	98.40%	39.20%	70.20%	92.00%	0.00%	88.45%	99.20%
Trial 9	44.00%	77.68%	98.40%	39.20%	71.01%	91.20%	0.00%	90.08%	99.20%
Trial 10	42.40%	76.62%	97.60%	39.20%	70.04%	92.00%	0.00%	89.78%	99.20%

Based on Table 7, from 10 trials of linear kernel optimization produce the lowest accuracy by 42.40% on 10th trial and produce the highest accuracy by 98.40% on 1st, 2nd, 3rd, 7th , 8th, and 9th trial. Polynomial kernels produce the lowest accuracy by 39.20% in every trials and the highest accuracy by 92% in 7th, 8th, and 10th trial. While RBF kernels produce the highest accuracy by 99.20% in every trials and the lowest accuracy by 0%. After conduct a checking, accuracy by 0% is affected by γ parameters with a value <=0.53. Example on parameter values that produce accuracy towards 0% is presented on Table 8.

TABLE 8
LOWEST ACCURACY SAMPLES IN RBF KERNEL

Trial-	Iteration-	genes-	C	γ	Accuracy
1	336	1	18.1866	0.6457	1.60%
1	802	1	43.3245	0.5829	0.80%
1	804	5	49.9636	0.5341	0.00%
3	622	3	10.8298	0.6471	1.60%
3	780	2	26.9130	0.5461	0.80%
4	393	8	21.0983	0.6190	1.60%
4	431	2	30.6210	0.5593	0.80%
4	370	2	7.0335	0.5491	0.80%
4	438	2	42.3367	0.4994	0.00%

Based on Table 8 can be seen that accuracy by 0% on RBF kernels occurs as a result from a value γ<=0.53. C parameters isn't too affecting accuracy. Accuracy by 0.8% obtained if it uses a value 0.5461<γ<0.6190. Furthermore, the greater the γ value will increase the accuracy until the highest accuracy by 99.20% with γ parameters range in between 6.6496 until 8.2911. Example on the highest accuracy on RBF kernels is presented on Table 9.

Table 9
HIGHEST ACCURACY SAMPLES IN RBF KERNEL

Trial-	Iteration-	genes-	C	γ	Accuracy
4	46	1	15.7499	6.8449	99.20%
4	56	1	14.1641	6.6496	99.20%
4	68	2	30.3772	7.9493	99.20%
4	185	2	44.0028	8.9150	99.20%
4	186	3	44.1127	8.9211	99.20%
4	794	3	26.0344	7.4171	99.20%
4	35	4	17.6879	7.1552	99.20%
4	180	4	24.4749	7.3605	99.20%
4	29	5	41.3318	8.7338	99.20%

SVM modeling using determined parameters and optimized parameters using genetic algorithm, certainly there is an increase in accuracy. Table 10 shows an increase in accuracy between determined parameters modeling and optimized parameters modeling.

Table 10
COMPARISON BEFORE AND AFTER OPTIMIZATION

	Before optimization	After optimization	Increment
Linear	83.20%	98.40%	15.20%
Polynomial	88.00%	92.00%	4%
RBF	98.40%	99.20%	0.8%

Based on Table 10, the improvement of linear kernel accuracy is by 15.20%, polynomial kernel by 4%, and RBF kernel by 0.8%. A linear kernel is a kernel that has the highest increased by 15,20%. This shows that linear kernel is the most unstable kernel if it given C value variations. On the contrary the RBF kernel is a SVM kernel which has the least increase in accuracy. This shows that RBF kernel is the most stable kernel in keeping the accuracy. RBF kernel average is always much higher compared to linear kernel and polynomial kernel. Testing the SVM parameters optimization using genetic algorithm reach the highest accuracy on 76th iteration as presented on Figure 4.

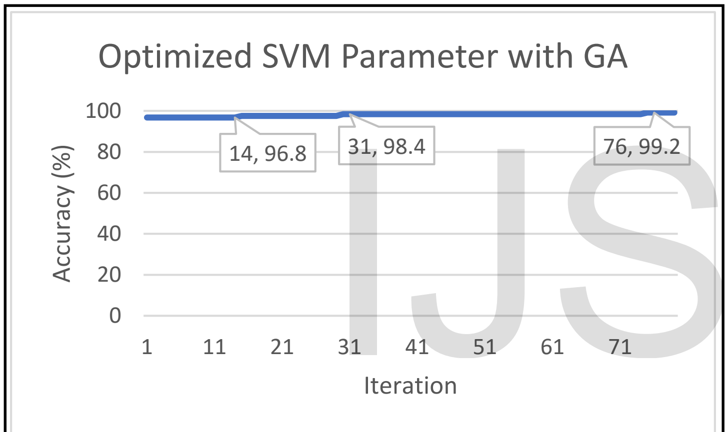

Fig. 4. Optimized SVM parameter with GA

3.7 Highest Accuracy Analysis based on Speakers

Based on Table 7 it can be seen that out of 10 repetitions, RBF kernel always produce the highest accuracy. If every voice being averaged as presented on Table 11.

Table 11
AVERAGE ACCURACY BY SPEAKER

	Average
Speaker-1	98.53%
Speaker-2	98.13%
Speaker-3	92.00%
Speaker-4	98.30%
Speaker-5	94.40%

Based on Table 6 it can be seen that the average accuracy on voice 3 looks the smallest compared with the other voice. After it being listened, it turns out the speaker for voice 3 has a little dialect. Different from speaker for voice 3, on speaker for voice 5 has the second lowest accuracy caused by noise such as whistling sound, activity sounds in the speaker environment, to the sound of pounding objects on the table.

4 CONCLUSION AND SUGGESTION

4.1 Conclusion

The conclusion of this study are SVM optimization using genetic algorithm can increase accuracy by 0.8% from the previous 98.40% into 99.20% obtained by RBF kernels. The fastest modeling time for data with 13x100 dimension too obtained by RBF kernels with a time of 0.68 seconds per 5 parameters value testing.

4.2 Suggestion

This research still have flaws that can be developed and fixed on the next research. Adding SVM kernels such as sigmoid kernel as accuracy comparison. Noise addition can be done to test how good RBF kernels keeping the accuracy.

ACKNOWLEDGMENT

The authors wish to thank Laboratory Computer Science FMIPA IPB Bogor by grant from used all facilities especially to collect all data.

REFERENCES

[1] Cortes C, Vapnik V. 1995. Support vector networks [Internet]. [downloaded on 2013 Mar 22]. Available on http://www.springerlink.com/index/K238JX04HM87J80G.pdf
[2] Goldberg, David E. 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. England : Addison-Wesley Publishing Company.
[3] Hsu C, Lin C. 2002. A comparison of methods for multiclass support vector machines. IEEE Transaction on Neural Networks. 13(2):415-425.
[4] Kamruzaman. 2010. Speaker Identification using MFCC Domain Support Vector Machine. Bangladesh. Rajshahi University.
[5] Kinnunen T. 2010. An Overview of Text-Independent Speaker Recognition from Features to Supervectors. Sciencedirect. Speech Communication 52 (2010) 12-40.
[6] Kohavi R. 1995. A research of cross validation and bootstrap for accuracy estimation and modem selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence. Quebec, Kanada. page 1137-1193.
[7] Lessman S, Stahlbock R, Crone S. 2006. Genetic Algorithm for Support Vector Machine. Canada. International Join Conference on Neural Network.
[8] Marinov. 2003. Text Dependent and Independent Speaker Verification Systems. Technology and Application. Paper on voice recognition technology. Downloaded on http://www.speech.kth.se/~rolf/gslt_papers/SvetoslavMarinov.pdf.
[9] Michalewicz Z. 1996. Genetic Algorithms and Data Structures = Evolution Programs. USA : Department of Computer Science, University of North Carolina.

[10] Pelton GE. 1993. Voice Processing. Singapore (SG): McGraw Hill.
[11] Shady M, Sharway M, Zayed H. 2014. Speaker Independent Arabic Speech Recognition Using Support Vector Machine. Mesir. Benha University.
[12] Smith N, Games M. 2012. Speech Recognition using SVM. Inggris: Cambridge University.
[13] Wang S, Mathew A, Chen Y, Ci L, Ma L, Lee J. 2009. Empirical analysis of support vector machine ensemble classifiers. Expert System with Applications. 36:6466-6476.