Implementation of support vector machine for classification of speech marked hijaiyah letters based on Mel frequency cepstrum coefficient feature extraction

To cite this article: Wisnu Adhi Pradana et al 2018 J. Phys.: Conf. Ser. 971 012050

View the article online for updates and enhancements.
Implementation of support vector machine for classification of speech marked hijaiyah letters based on Mel frequency cepstrum coefficient feature extraction

Wisnu Adhi Pradana, Adiwijaya, Untari Novia Wisesty
School of Computing, Telkom University, Bandung 40257, Indonesia

wa.pradana01@gmail.com, adiwijaya@telkomuniversitiy.ac.id, untarinw@telkomuniversitiy.ac.id

Abstract. Support Vector Machine or commonly called SVM is one method that can be used to process the classification of a data. SVM classifies data from 2 different classes with hyperplane. In this study, the system was built using SVM to develop Arabic Speech Recognition. In the development of the system, there are 2 kinds of speakers that have been tested that is dependent speakers and independent speakers. The results from this system is an accuracy of 85.32% for speaker dependent and 61.16% for independent speakers.

1. Introduction
Arabic is one of the oldest languages in the world, there is little Speech Recognition in Arabic than any other language. In its development, Arabic Speech Recognition has many problems that make it difficult to develop. The most influencing factor in developing the Speech Recognition Arabic language is the large variety of dialect and morphological complexity [2]. Dialect variations can cause meaning differences in spoken word. In addition, in arabic there are some rules on how to read the word (tajwid) so it can affect the meaning of a word.

Speech Recognition or abbreviated SR is a technology that enables computers to recognize words spoken by humans on a mic or phone. In its use, SR has been used in a variety of problems, such as the use of computer command with voice or learning a foreign language. In addition, SR also helps interact with foreign people.

There are many methods that have been used in building an SR. However, the method often used to build an SR is the Hidden Markov Model or commonly called HMM. HMM is a model that assumes parameters of unknown parameters and hidden parameters. Therefore, the Speech Recognition Arabic that has evolved today is using the HMM approach [6].

Support Vector Machine or commonly called SVM is one method that can be used to process the classification of a data [4]. The author intends to use the approach in the development of Arabic Speech Recognition. With this system, it is expected to increase the choice in using Arabic Speech Recognition as a classifier of hijaiyah marked letters.

In this paper, the result of implementation Support Vector Machine for Arabic Speech Recognition is given. The proses of Mel Frequency Cepstrum Coefficient will be elaborated in section 2. In the next
section. Support Vector Machine proses will be explained. In the section 4, the result has been given to in detail. In the last section, there are the conclusion for all paper.

2. Mel Frequency Cepstrum Coefficient
Mel Frequency Cepstrum Coefficient or MFCC is the most commonly used approach for speech recognition. MFCC is the only acoustic approach that takes the human perception (physiology and behavioral aspects of the sound production organs) as a consideration for building speech recognition which means this approach can not process frequencies above 1KHz. This approach has two types of filters that are linearly at low frequencies below 1000Hz and logarithmic distances above 1000Hz.

There are several steps in the MFCC algorithm to generate a feature vector. These steps are pre-emphasis, frame blocking, windowing, Fast Fourier Transform, Mel Filter Bank Processing, discrete cosine transformation and Delta Energy and Delta Spectrum. The result of MFCC is a feature vector that will be used as a database to determine the classification of voice data.

![Figure 1. The Step of MFCC](image-url)

3. Support vector machine
Support vector machine or can be called SVM is one method of supervised learning that can be used for data classification [4]. SVM divides the data into two classes by maximizing margins between hyperplane with the two closest data in both classes [4]. The most optimal hyperplane is called a support vector. To obtain the best hyperplane calculations are needed to determine the distance between data with existing hyperplane. In the first learning process, we will find the minimum distance from each training data on hyperplane using formula 1 and 2.

\[
y_i(w^T, x_i + b) \geq 1 \quad ; i = 1,2 \ldots n
\]

\[x_i = \text{data input}
\]

\[y_i = \text{target}
\]

\[b = \text{position of the plane towards the coordinate center}
\]

\[w = \text{normal plane}
\]

Equation 2 can be solved using the optimal point of the lagrange function.

\[
L_p = L(w,b,a) = \frac{1}{2} ||w||^2 - \sum_{i=1}^{n} \alpha_i(y_i(w^Tx_i+b)-1)
\]

\[\alpha_i \text{is a lagrange multiplier. To find the optimal point, then w and b must be minimized so that L = 0. Under these conditions, then } \alpha \text{ can be known.}
\]

To make a decision from data testing, it takes a decision function created from the kernel function. There are many kernel functions that can be used for decision making, here are examples of frequently used functions:
Linear kernel: \( K(x_i, x_j) = x_i^T x_j \) \hspace{1cm} (3)

Polynomial kernel: \( K(x_i, x_j) = (\gamma x_i^T x_j + r)^d \) \hspace{1cm} (4)

\( d = \) degree  \\
\( x_i \text{ dan } x_j = \) vector data  \\
\( \gamma \text{ dan } r = \) parameter

Quadratic kernels also include polynomial kernels, but with a second order [16].

3.1. Multi-class SVM

Initially SVM can only be used to classify 2 classes only. However, in its development the researchers have developed a method that makes SVM can be used for the classification of more than 2 data classes. There is a way to implement multi-class SVM by combining some binary SVM or combining all data into an optimization system.

The "one against all" method is one of the methods that can be used for multi-class SVM problems. This method builds n the binary SVM model (n is the number of classes). Each model is in-training with all the data to look for solution. See table 2.1 and figure 3.2

| \( \gamma \) | \( y_i = 1 \) | \( y_i = -1 \) | Hypothesis |
|---|---|---|---|
| Class 1 | Not class 1 | \( f_1(x) = (w_1)x + b_1 \) |
| Class 2 | Not class 2 | \( f_2(x) = (w_2)x + b_2 \) |
| Class 3 | Not class 3 | \( f_3(x) = (w_3)x + b_3 \) |
| Class 4 | Not class 4 | \( f_4(x) = (w_4)x + b_4 \) |

4. Result and Comparison

4.1. Dependent speaker

Accuracy is a measure that approximates the correct size or is received from the quantity amount being measured. Accuracy gained from the number of test results is true divided by the amount of tested data. Table 4.1 is given to represent the results of the types of dependent speakers that have been produced during the study.
Table 4.1 illustrates the accuracy obtained on each letter that has been tested against the model with 2 different kernel functions. The biggest accuracy obtained by using linear and quadratic kernel functions is \( \text{ro} \) with 97.22% for linear kernel function and 61.11% for quadratic kernel function. The smallest accuracy obtained by using linear kernel function and quadratic kernel function is \( \text{tsa} \) with 72.22% for linear kernel function and 22.22% for quadratic kernel function. This indicates that the optimized model is used to classify \( \text{ro} \) letters rather than \( \text{tsa} \). In addition, the average accuracy obtained from systems with linear kernel function is 85.32% and quadratic kernel function is 43.45%. The distribution of letters that run on the system also affects the results of accuracy. With this division, the system does not need to create 168 class models directly. In addition the process of making the model can diroleh more quickly.

### Table 2. result of testing data with dependent speaker

| No | Letter | Hijaiyah | Accuracy Linear | Accuracy Quadratic | No | Letter | Hijaiyah | Accuracy Linear | Accuracy Quadratic |
|----|--------|----------|----------------|-------------------|----|--------|----------|----------------|-------------------|
| 1  | Alif   | ا        | 88.89%         | 44.44%            | 15 | Dhod  | ص         | 88.89%         | 47.22%            |
| 2  | Ba     | ب        | 75%            | 33.33%            | 16 | Tho   | ط         | 91.67%         | 52.78%            |
| 3  | Ta     | ت        | 77.78%         | 38.89%            | 17 | Zho   | ظ         | 80.56%         | 44.44%            |
| 4  | Tsá    | ث        | 72.22%         | 22.22%            | 18 | Ain   | غ         | 86.11%         | 58.33%            |
| 5  | Jim    | ج        | 86.11%         | 27.78%            | 19 | Ghoin | غ         | 77.78%         | 52.78%            |
| 6  | Ha     | ح        | 88.89%         | 52.78%            | 20 | Fa    | ف         | 83.33%         | 33.33%            |
| 7  | Kho    | خ        | 94.44%         | 44.44%            | 21 | Qof   | ق         | 86.11%         | 44.44%            |
| 8  | Dal    | د        | 88.89%         | 38.89%            | 22 | Kaf   | ك         | 86.11%         | 38.89%            |
| 9  | Dzal   | ذ        | 86.11%         | 41.67%            | 23 | Lam   | ل         | 88.89%         | 47.22%            |
| 10 | Ro     | ر        | 97.22%         | 61.11%            | 24 | Mim   | م         | 86.11%         | 30.56%            |
| 11 | Zai    | ز        | 88.89%         | 50%               | 25 | Nun   | ن         | 80.56%         | 47.22%            |
| 12 | Sin    | س        | 94.44%         | 47.22%            | 26 | Wa    | و         | 80.56%         | 41.67%            |
| 13 | Syin   | ش        | 80.56%         | 36.11%            | 27 | Ha    | ه         | 88.89%         | 44.44%            |
| 14 | Shat   | ص        | 83.33%         | 38.89%            | 28 | Ya    | ي         | 80.56%         | 55.56%            |

#### 4.2. Independent speaker

Tests on independent speaker types are performed using linear kernel functions. Table 4.2 represents the results of testing using linear kernel functions. There are 672 of testing data that has been tested in this system. This data comes from 1 person who speaks all the letters with different harakat.

In table 4.2 can be seen that the biggest accuracy obtained the letters \( \text{sin} \) (س) and \( \text{tho} \) (ط) is 87.5%. While the smallest accuracy obtained from the test results of this model is the letter \( \text{fa} \) (ف) with an accuracy of 25%. This indicates that the model is more optimize to classify the letters \( \text{sin} \) and \( \text{tho} \) rather than the letter \( \text{fa} \). In addition, the average accuracy obtained from the system is 61.16%.

### Table 3. result of testing data with independent speaker

| No | Letter | Hijaiyah | Accuracy | No | Letter | Hijaiyah | Accuracy |
|----|--------|----------|----------|----|--------|----------|----------|
| 1  | Alif   | ا        | 75%      | 15 | Dhod  | ص         | 75%      |
| 2  | Ba     | ب        | 54.17%   | 16 | Tho   | ط         | 87.50%   |
| 3  | Ta     | ت        | 75%      | 17 | Zho   | ظ         | 83.33%   |
| 4  | Tsá    | ث        | 54.17%   | 18 | Ain   | ع         | 75%      |
| 5  | Jim    | ج        | 66.67%   | 19 | Ghoin | غ         | 41.67%   |
| 6  | Ha     | ح        | 41.67%   | 20 | Fa    | ف         | 25%      |
4.3. Comparison
The best accuracy that obtained from support vector machine model with linear kernel function is \(ro (\text{ر})\) with 97.22% accuracy value for dependent speaker data testing. While for independent speaker test data, the best accuracy result obtained from support vector machine model with linear kernel function is \(sin (\text{س})\) and \(tho (\text{ث})\) with 87.5% accuracy. In addition, the average accuracy of both types is 85.32% for dependent speakers and 61.16% for independent speakers.

Meanwhile, support vector machine model with quadratic kernel function is only obtained from dependent speaker type. This happens because the support vector machine model with quadratic kernel function fails to classify data testing from independent speakers. The best accuracy results obtained from the support vector machine model with quadratic kernel function is the letter of \(ro (\text{ر})\) with an accuracy value of 61.11%. In addition, the average accuracy obtained from this model is 43.45%.

From the results of tests that have been done that the support vector machine model with the type of dependent speaker is better than the independent speaker type. This proves the dependency of a data affecting the results of the classification. In addition, kernel function also affects the result of classification, it is evident that linear function is better than quadratic function.

5. Conclusion
Based on the results of tests that have been implemented, it can be drawn some conclusions as follows, the treatment of training and testing into each letter (each letter has 6 classes) makes the system work faster than all data is processed together. The kernel function affects the accuracy of the system with the use of linear functions resulting in an average accuracy of 85.32% for speaker dependent and 61.16% for independent speakers. Dependency of a data affects the classification results. Models that use dependent speakers have better accuracy than independent speakers.

References
[1] Abushariah, M.A.A.M., Ainon, R., Zainuddin, R., Elshafei, M. and Khalifa, O.O., 2012. Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus. International Arab Journal of Information Technology (IAJIT), 9(1), pp.84-93.
[2] AbuZeina, D. and Elshafei, M., 2012. Arabic Speech Recognition Systems. In Cross-Word Modeling for Arabic Speech Recognition (pp. 17-23). Springer US.
[3] Alotaibi, Y.A. and Hussain, A., 2010. Comparative analysis of arabic vowels using formants and an automatic speech recognition system.
[4] Durgesh, K.S. and Lekha, B., 2010. Data classification using support vector machine. Journal of Theoretical and Applied Information Technology, 12(1), pp.1-7.
[5] El Choubassi, M.M., El Khoury, H.E., Alagha, C.J., Skaf, J.A. and Al, M.A., 2003. Arabic speech recognition using recurrent neural networks. In IEEE Intl. Symp. Signal Processing and Information Technology ISSPIT.
[6] Juang, B.H. and Rabiner, L.R., 1991. Hidden Markov models for speech recognition. Technometrics, 33(3), pp.251-272.
[7] Kalinli, A. and Sagiroglu, S., 2006. Short Paper_. Journal of Information Science and
Engineering, 22, pp.1555-1568.

[8] Meng, Y.U.A.N., 2004. Speech recognition on DSP: Algorithm optimization and performance analysis (Doctoral dissertation, The Chinese University of Hong Kong).

[9] Muda, L., Begam, M. and Elamvazuthi, I., 2010. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083.

[10] Satori, H., Hiyassat, H., Harti, M. and Chenfour, N., 2009. Investigation arabic speech recognition using CMU sphinx system. Int. Arab J. Inf. Technol., 6(2), pp.186-190.

[11] Singh, S. and Rajan, E.G., 2011. Vector quantization approach for speaker recognition using MFCC and inverted MFCC. International Journal of Computer Applications, 17(1), pp.1-7.

[12] Shady, Y. and Zayed, S.H.H., 2009. Speaker independent Arabic speech recognition using support vector machine.

[13] Susanto, C.P. and Setiawan, E.I., 2015. Algoritma Support Vector Machine Untuk Mendeteksi Smst Spam Berbahasa Indonesia.

[14] U.N. Wisesty, M.S. Mubarok, Adiwijaya, A classification of marked hijaiyah letters’ pronunciation using hidden Markov model, AIP Conference Proceedings 1867(1), 020036

[15] Zhang, S.X., 2014. Structured support vector machines for speech recognition.

[16] Chang, Y.W., Hsieh, C.J., Chang, K.W., Ringgaard, M. and Lin, C.J., 2010. Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research, 11(Apr), pp.1471-1490.