ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.10, 8675–8683
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.10.93

TAYLOR DIRICHLET PROCESS MIXTURE FOR SPEECH PDF ESTIMATION AND SPEECH RECOGNITION

T. RAJESH KUMAR[1], C. M. VELU, C. KARTHIKEYAN, S. SIVAKUMAR, SREERAM NIMMAGADDA, AND D. HARITHA

ABSTRACT. For a backward Markova particle filter, the estimation of noise sequences and a strategy for modelling of speech probability distribution function using Taylor hybrid Dirichlet Process Mixture are deliberated in this manuscript. Generally, the speech identification rate, drastically reduces due to the effect of circumstance noises, echo signal, white noise and so on. The improvement of speech identification rate is required to increase by estimating the sequence of noises. In this manuscript, instead of using a traditional methods of Gaussian Mixture Model (GMM), the clear speech is modelled using Taylor-DPM method. With the help of parameters such as standard deviation, covariance and mean created using Taylor-DPM, a sequence of speech signal will be generated. This method uses Kalman filter to improve the identification of speech in Signal to Noise ratio and estimates the noise signal sequences using backward Markova particle filter.

1. INTRODUCTION

General1y, Scientific research in engineering, science and technology deals with mathematical formulae, functions and models. For certain application in scientific engineering field such as image/speech/signal processing, Dirichlet process provides a technical support in all aspects. The journal and tutorial

[1]*corresponding author*

2020 *Mathematics Subject Classification.* 05C78, 05C05.

Key words and phrases. Signal processing, Speech detection, Stochastic Taylor hybrid Dirichlet Process Mixture, Distributed Kalman filter, Backward Markov particle filter.

papers[1,2] discussed the importance of Dirichlet process is used in this manuscript for speech density estimation speech detection. The manuscripts [3-5] deliberates the importance of particle filters and their applications in speech field, hand writing recognition. Without the utilization of traditional method, [6,7] the probability distribution function of speech is modelled using (DPM) Dirichlet Process Mixture. A non constraints probability distribution function, distributes over, all possible space is defined as Dirichlet Process. Without the intrusion of GMM unwanted signals probability density function, some papers shows the enhancement in paradigm algorithm [9,10]. The TIMIT database is used as a bench mark. The non-audible murmur or whispered voice recognition and identification using inverse filters also discussed in some manuscripts [11-16]. The language English and other regional language of Telugu, speech emotion recognized using the MFCC hybrid with Hidden Markov Model(HMM) also discussed in manuscript of [17-20]. But this manuscript deals with mathematically integrating Taylor function coefficients with Dirichlet process mixture for identification of speech signal. The data set used is TIMIT corpus.

2. Proposed methodology

Without using GMM or FR-GMM model a clear speech modelling is proposed using Taylor hybrid Dirichlet Process Mixture for speech signal identification or recognition. This method uses the protocol of Gaussian and Markov, since we add the particle filter as on of the components in noise estimation.

2.1. **Mathematically added noise in clear speech signal.**

When a clear speech signal is added with noise, the additive noise is elaborated with the help of Gaussian probability distribution function [7]. Then, the statistical model is

$$p(x) = \sum_{k=1}^{K} P(v_k)\Re(x, \mu_{x,k}\sigma_{x,k}).$$

Here, $\Re(x, \mu, \sigma$ be the Gaussian distribution with $p(x)$ as probability function. Let the input clear speech signal be S_T, the additive noise signal be n_T, also auxiliary independent constants such as u_T, w_T are added to the signal, then the resultant x_T be as follows:

$$x_T = S_T + log(1 + exp(n_T - S_T)) + u_T,$$

$$n_T = n_{T-1} + w_{T-1},$$

$$u_T \cong \Re(0, \sigma_S) \ \& \ w_T \cong \Re(0, \sigma_w).$$

2.2. **Probability Function Distribution Estimation and Dirichlet Processes Mixture.** Let the sound sequence be a statistical function of probability density function, then $y_1, y_2, \ldots, y_n$ with non constraint model estimates the $K(.)$ as:

$$K(y) = \int_{\theta} k(y/\theta)\, dH(\theta), \tag{2.1}$$

where θ is a group variable of sound, $k(y/\theta)$ and $H(\theta)$ are the Bayesian framework with mixed pdf having random probability measure [8]. Let the Dirichlet process continues. The probability quantify H_0 on space on any framework of speech be $'B'$ as $B_1, B_2, \ldots, B_t$ and $'\alpha'$ be the +ve integer number, then a distribution of probability $'H'$, as per Dirichlet process the $H \cong DP(H_0, \alpha)$, in which $'D'$ be the Diritchlet distribution at standard mode

$$(H(B_1), H(B_2), \ldots, H(B_t)) \cong D(H_0(B_1), H_0(B_2), \ldots, H_0(B_t), \alpha). \tag{2.2}$$

The poly urn representation of Dirichlet process distribution over clear speech with added noise using Bayesian framework is given as equation (2.2) is integrated with (2.1):

$$\theta_{t+1}|\theta_t \approx \frac{1}{t+\alpha} \sum_{k=1}^{t} \delta_{\theta_k} + \frac{\alpha}{t+\alpha} H_0,$$

wheres δ_{θ_k} be the delta function.

Now the mixture of Dirichlet process with random probability quantifier and input clear speech signal is modelled to estimate the density problem with the help of DPM,

$$H \cong DP(H_0, \alpha);\ \theta_k \cong H;\ \ S_T \cong f(S|\theta_k).$$

The probability distribution which are not knowing is followed by simple forthcoming function equation:

$$G(S) = \int_{\Theta} f(S|\theta) dH(\theta) \quad with\ \theta \in \Theta.$$

2.3. **Taylor Series-Coefficient hybrid DPM.** As per the standard Taylor equation shown in [11,12], the speech features frames are mingled with the Taylor series is given as,

$$\begin{aligned} H(t+1) =& 0.5H(t) + 1.3591H(t-1) - 1.3590H(t-2) \\ &+ 0.6795H(t-3) - 0.2259H(t-4) + 0.055H(t-5) \\ &- 0.010H(t-6) + 1.38e^{-3}H(t-7) - 9.92e^{-5})H(t-8). \end{aligned} \tag{2.3}$$

The advantage of Taylor series is that it ensures the accurate estimation of the particle filter coefficients and it attains convergence easily. The equation (2.3) is modified as follows,

$$\begin{aligned} H(t+1) =& 0.5H(t-1) + 1.3591H(t-2) - 1.3590H(t-3) \\ &+ 0.6795H(t-4) - 0.225H(t-5) + 0.055H(t-6) \\ &- 0.010H(t-7) + 1.38e^{-3}H(t-8) - 9.92e^{-5}H(t-9). \end{aligned}$$

Rearranging the above equation, we get,

$$H(t-1) = 2\begin{bmatrix} H(t) - 1.3591H(t-2) + 1.3590H(t-3) \\ -0.6795H(t-4) + 0.2259H(t-5) - 0.055H(t-6) \\ +0.010H(t-7) - 1.38e^{-3}H(t-8) + 9.92e^{-5}H(t-9) \end{bmatrix},$$

$$H(t-1) = \begin{bmatrix} 2H(t) - 2.7182H(t-2) + 2.718H(t-3) \\ -1.359H(t-4) + 0.4518H(t-5) - 0.111H(t-6) \\ +0.0208H(t-7) - 0.00276e^{-3}H(t-8) \\ +0.00019H(t-9) \end{bmatrix}. \tag{2.4}$$

The novel Taylor coefficients are extracted by substituting the equation (2.4) in Dirichlet Process Mixture algorithm that is given below.

$$H(t, G(S)) = \Delta H_T(t, G(S)) + \begin{bmatrix} 2H(t) - 2.7182H(t-2) \\ +2.718H(t-3) - 1.359H(t-4) \\ +0.4518H(t-5) - 0.111H(t-6) \\ +0.0208H(t-7) - 0.00276e^{-3} \\ H(t-8) + 0.00019H(t-9) \end{bmatrix}. \tag{2.5}$$

The above equation (2.1) is Taylor hybrid DPM. The Algorithm starts with the population of Taylor-DPM being initialized. With every iteration, the Taylor-DPM travels from local to global solutions. If a Dirichlet process mixture seeks

a better solution after jumping. The process is continued until to the end conditions are met. This solution is the best solution finally.

2.4. **Statistical Taylor-DPM for the evaluation of speech signal PFD.** With the available coefficients, the evolution of noise and clear speech sequences are to be generated using the following probability: $p(n_{0:k}, \theta_{1:k}|x_{1:k}) = p(n_{0:k}|\theta_{1:k}, x_{1:k})$ $p(\theta_{1:k}|x_{1:k})$ in which mean vector and standard deviation matrix of clear speech signals μ_t and σ_k based on θ_k. The $H_0 \cong \Re IW(\mu_0, k_0, u_0\sigma_0)$ indicates the Wishart's inverse distribution are the hyper parameters of distributions.

The probability may be calculated using Extended Kalman Filter (EKF) using particle filter implementation method as per given in Kenko Ota et. al. manuscripts. Now apply the particle filter from Kalman as "J" particles provides the probability as:

$$P_N(n_k, \theta_{1:k}) = \sum_{i=1}^{J} \hat{\omega}_t^i \; p(n_k|\theta_{1:k}^i, x_{1:k}$$

with

$$p(n_k|\theta_{1:k}^i, x_{1:k} \cong \Re(\hat{n}_{k|k}(\theta_{1:k}^i), \sigma_{n_{k|k}}^i(\theta_{1:k}^i))$$

for each 'i' value of Extended Kalman filter particle, the recursive computation happens for $(\hat{n}_{k|k}(\theta_{1:k}^i) \;\; and \; \sigma_{n_{k|k}}^i(\theta_{1:k}^i))$ parameters. Thus by using weights, the estimation of probability and the coefficients are calculated. Finally we get an estimation of : $\hat{\omega}_k^j \propto \omega_{k-1}^i \Re(\hat{x}_k(\theta_{1:k}^i), \sigma_x(\theta_{1:k}^i))$ Based on the polya urn depiction, we identify the $p(\theta_k^i|\theta_{k-1}^i)$ and coefficients for the speech signal.

2.5. **Identification of speech/speech less frames using mathematical Model.** The manuscript identifies the maximum Signal to Noise ratio (SNR) to differentiate the speech and speech less portion of the a frame of speech. The distance of speech specifies the identification of speech,

$$\text{Signal: } Dist_{s_k} = (x_k - (s_k + log(1 + exp(n_k - \hat{s}_k))))^2$$

$$\text{Noise: } Dist_{n_k} = (x_k - \hat{n}_k)^2$$

$$\text{Difference: } \Delta \, diff_t = Dist_{\,S_k} - Dist_{n_k}$$

Whereas noise and clear speech are $\hat{n}_k$ and $\hat{S}_k$ are respectively estimated and will be modified as follows with Taylor wiener coefficient constants:

$$\hat{S}_k = \hat{S}_k + \psi_S \sqrt{Dist_{S_k}};\; \hat{n}_k - \psi_n \sqrt{Dist_{S_k}}$$

Finally the proposed Taylor hybrid Dirichlet Process Mixture identifies the speech and its pdf distribution with the help of the coefficients identified by Ψ_S and Ψ_n. These steps may be written as an algorithm.

3. Data set, results and discussions

The phonemically and lexically facilitated corpus available in TIMIT corpus [10] speech of English words are uttered with 16 kHz and 16-bit speech waveform. In this experimental setup with clear speech, noises of white and particle noise are synthetically added at a Signal to Noise Ratio of 0db to 8db. For the evaluation purpose 108 words are uttered by 2 males and 1 female are used. This manuscript compares three types of data: 1) no processing on signal (NIL Filter) 2)Expanded Kalman filter and 3) Traditional GMM processed filters. The following table, Table 1 shows the Speech identification rate for percentage of noise data for different filters used.

Table 1. Speech Identification using filters for % of Noise added.

Speech in decibels	White			Particle		
	Nil Filter	GMM	EKF	Nil Filter	GMM	EKF
0db	2.8	3.1	22.5	7.2	10.2	29.5
2db	16.2	10.9	45.2	28.4	19.2	57.2
4db	48.7	32.7	68.4	57.3	38.6	69.7
8db	79.2	49.3	82.6	81.2	54.1	86.3

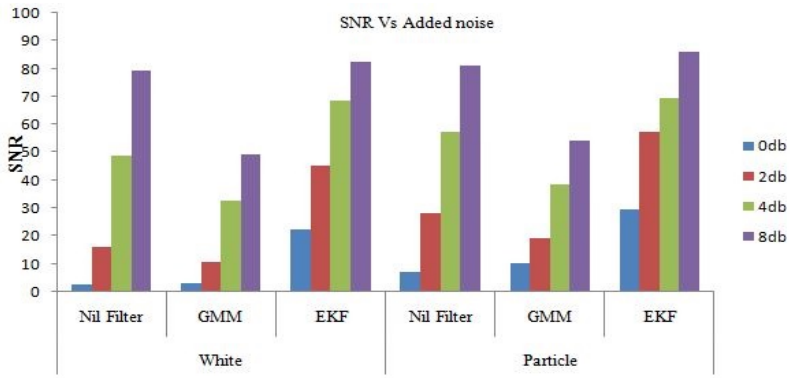

Figure 1. Speech Identification using SNR Results

The Figure 1. stipulates the speech identification using filters for percentage of noise added, in which SNR vs added noise with different filters are shown.

4. Conclusion

A technique for modelling of clear speech probability density function using Taylor-Dirichlet Processing Mixture for identification of speech, in which a sequence of particle noise added filter is discussed in this manuscript. Also how to integrate the Taylor series with DPM is highlighted. Rather than GMM traditional method having incorrect preciseness, this technique proves the good estimation of noise, accurately. While evaluating the SNR's result for speech identification, the particle noise added with extended Kalman filter provides the improved result is experimentally verified.

References

[1] Y. Li, E. Schofield, M. Gonen: *A tutorial on Dirichlet process mixture modelling*, Journal of Mathematical Psychology, Elsevier, **91** (2019), 128-144.

[2] J. Paisley: *A Tutorial on the Dirichlet Process for Engineers Technical Report*, Department of Electrical & Computer Engineering Duke University, Durham, NC, (2018), 1-23.

[3] A. Doucet, A. M. Johansen: *A Tutorial on Particle Filtering and Smoothing*, The Institute of Statistical Mathematics, Japan, Department of Statistics, (2012), 1- 41.

[4] S. Sivakumar, S. R. Nayak, A. Kumar, S. Vidyanandini: An empirical study of supervised learning methods for Breast Cancer Diseases, Optics, **175** (2018), 105-114.

[5] M. D. Escobar, M. West: *Bayesian Density Estimation and Inference using Mixtures*, Journal of the American Statistical Association, **90**(430) (1995), 577-588.

[6] A. Kottas: *Dirichlet Process Mixtures of Beta Distributions, with Applications to Density and Intensity Estimation*, Proceedings of Nonparametric Bayesian Methods, 23rd ICML, 2006.

[7] K. Ota, E. Duflos, P. Vanheeghe, M. Yanagida: *Speech Recognition and Speech Density Estimation by the Dirichlet Process Mixture*, Proceedings of ICASSP 2008, IEEE xprorer, (2008), 1553-1556.

[8] F. Caron, M. Davy, A. Doucet, E. Duflos, P. Vanheeghe: *Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures*, Proc. of Fusion'06, Florence, Italia, (2006), 10-13.

[9] J. C. Segura, A. de la Torre, M. C. Benitez, A. M. Peinado: *Model-Based Compensation of the Additive Noise for Continuous Speech Recognition Experiments Using AURORA II Database and Tasks*, Proc. EuroSpeech '01(I), (2001), 221-224.

[10] TIMIT ACOUSTIC-PHONETIC CONTINUOUS SPEECH CORPUS: 2018. https:\\catalog.ldc.upenn.edu\.

[11] P. KURADA, S. MARUVADA, K. R. SANAGAPALLEA: *Speech bandwidth extension using transform-domain data hiding*, International Journal of Speech Technology, **22**(2) (2019), 23-32.

[12] T. RAJESH KUMAR, G. R. SURESH, S. KANAGA SUBARAJA, C. KARTHIKEYAN: *Taylor-AMS features and deep convolutional neural network for converting non-audible murmur to normal speech*, Computational Intelligence, (2020), 1-24.

[13] M. N. SHARIFF, B. SAISAMBASIVARAO, T. VISHVAK, T. RAJESH KUMAR: *Biometric user identity verification using speech recognition based on ANN/HMM*, Journal of Advanced Research in Dynamical and Control Systems, **9**(12) (2017), 1739-1748.

[14] T. RAJESH KUMAR, M. SRINAGAMANI, M. SAI RAM CHANDU, S. MOUNIKA: *Conversion of body conducted unvoiced speech(murmur) to normal speech using Hidden Markov Model (HMM)*, International Journal of Engineering and Technology(UAE), **7**(2.32) (2018), 400-404.

[15] T. RAJESH KUMAR, G. R. SURESH, S. KANAGA SUBA RAJA: *Conversion of Non Audible Murmur to Normal Speech based on Full-rank Gaussian Mixture Model*, Journal of Computational and Theoretical NanoScience, **1**(15), (2018), 185-190.

[16] T. RAJESH KUMAR, G. R. SURESH, K. S. SUBARAJA: *Conversion of non-audible murmur to normal speech based on GR-GMM using Non-Parallal Training Adaptation Method*, Proceesdings of ICISS-2019, IEEE Xplore, 2019.

[17] K. MANNEPALLI, P. N. SASTRY, M. SUMAN: *Accent recognition system using deep belief networks for telugu speech signals*, Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, 2017, 99-105.

[18] J. GIRIKA, M. Z. UR RAHMAN: *Adaptive speech enhancement techniques for computer based speaker recognition*, Journal of Theoretical and Applied Information Technology, **95**(10) (2017), 2214-2223.

[19] V. RAGHAVARAO, Y. SASIDHAR, K. SASTRY, J. K. R. CHANDRA PRAKASH: *Quantifying quality of WEB sites based on content*, International Journal of Engineering and Technology (UAE), **7**(2) (2018), 138-141.

[20] K. MANNEPALLI, P. N. SASTRY: *Accent recognition system using deep belief networks for Telugu speech signals*, International Journal of Speech Technology, **19**(1), (2017), 87-93.

Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh - 522502.
Email address: t.rajesh61074@kluniversity.in

Department of Computer Science and Engineering
Saveetha School of Engineering, Chennai, Tamil Nadu

Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh - 522502.

Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh - 522502.

Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh - 522502.

Department of Computer Science and Engineering
1Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh - 522502.