논문 2009-46SP-3-17

최소 분류 오차 기법을 이용한 보이스 피싱 검출 알고리즘

(Voice-Pishing Detection Algorithm Based on Minimum Classification Error Technique)

이 계 환*, 장 준 혁**

(Kye-Hwan Lee and Joon-Hyuk Chang)

요 약

본 논문에서는 보이스 피싱 (Voice Pishing) 예방을 위한 알고리즘을 최소 분류 오차 기법 (Minimum Classification Error)을 기반으로 제한하다. 휴대폰으로 전송되어진 신호를 기반으로 3GPP2 Selectable Mode Vocoder (SMV)의 복호화 과정에서 자동적으로 추출되는 중요 특징벡터를 사용하여 Gaussian Mixture Model (GMM)을 구성하고 이를 기반으로 구해지는 로그 (Log) 기반의 우도 (Likelihood)를 사용한 변별적 가중치 학습을 사용하여 보이스 피싱 예방을 위한 검출 알고리즘을 제안하다. 실험 결과 제안된 보이스 피싱 알고리즘이 기존의 방법에 비해 우수한 성능을 보인 것을 알 수 있었다.

Abstract

We propose an effective voice-phishing detection algorithm based on discriminative weight training. The detection of voice phishing is performed based on a Gaussian mixture model (GMM) incorporaiting minimum classification error (MCE) technique. Actually, the MCE technique is based on log-likelihood from the decoding parameter of the SMV (Selectable Mode Vocoder) directly extracted from the decoding process in the mobile phone. According to the experimental result, the proposed approach is found to be effective for the voice phishing detection.

Keywords : 보이스 피싱, Selectable Mode Vocoder (SMV), Gaussian Mixture Model (GMM), Minimum Classification Error (MCE)

I. 서 론

일반적으로 보이스 피싱은 금전상의 이득을 목적으로 하여 대중으로 부터 그에 상응 하는 개인적 정보나 금융관련 정보 등을 얻기 위해 행해진 범죄 행동이다. 이러한 보이스 피싱은 금전상의 피해뿐만 아니라 개인 식별 번호 (Personal Identification Number, PIN), 유효기간, 생일 등과 같은 추가적인 정보를 목적으로 사용되어 지기도 한다. 또한, 일반적으로 보이스 피싱은 모니터링 하거나 추적하는 것이 매우 어렵기 때문에 현재 보이스 피싱을 막기 위해서는 금융관련 정보를 요구하거나 이와 유사한 행위가 요구될 때 소비자에게 강한 의심을 가지라는 충고를 하는 것이 유일한 실정이다.

이러한 보이스 피싱에 대한 보호를 위해 많은 연구들이 진행되고 있으며, 특히 사람의 잠재의식 속의 행동들에 기반한 연구가 많이 진행되고 있다. 보이스 피싱과 관련하여 사람의 잠재의식과 관련된 행동에 대해 많이 알려진 사실 중 하나는 사람들이 거짓말을 할 경우에 자신도 알기 힘든 변화가 있다는 것이다[1]. 사람들이 거짓말을 할 경우 목소리의 떨림, 눈동자의 움직임, 손동작, 얼굴의 미세한 표정변화 등과 현상들이 관찰된다. 이와 같은 현상은 비언어적 누출 (Nonverbal Leakage)이라는 표현으로 명명되어진다. 이러한 비언어적 누출

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Department of Electronics Engineering, Inha University)

※ 본 연구는 지식경제부 및 정보통신연구진흥원의 IT 핵심기술개발사업 [2008-F-045-01]과 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2008-C1090-0902-0010).

접수일자: 2008년9월9일, 수정완료일: 2009년4월13일

중에서 음성 정보는 거짓말을 할 때 이를 인지할 수 있게 하는 가장 중요한 정보 중 하나이다[2].

본 논문에서는 핸드폰을 이용한 전화사기를 예방하기 위하여 변별적 가중치 학습에 기반한 보이스 피싱 검출 알고리즘을 제안하다. 기존에 우리의 연구에서 보인 효과적인 특징벡터를 사용한 보이스 피싱 검출 알고리즘을 보다 향상시키기 위해 Minimum Classification Error (MCE)를 사용하여 구해진 변별적 가중치를 적용하였다[3~4]. 그 결과 제안된 방법이 기존의 방법보다 우수한 성능을 보임을 알 수 있었다.

본 논문의 구성으로는, Ⅱ장에서는 실험에서 사용되어진 Selectable Mode Vocoder (SMV)와 추출되어진 특징벡터에 대해서 기술하고, Ⅲ장에서는 제안된 변별적 가중치가 적용된 보이스 피싱 검출 알고리즘에 대해서 기술한다. Ⅳ장에서는 실험 결과에 대한 비교 및 분석에 대해 기술하였으며, 마지막으로 Ⅴ장에서 결론을 맺는다.

Ⅱ. SMV의 이해와 사용되어진 특징벡터

본 논문에서 사용되어진 음성 부호화기인 SMV는 3GPP2의 표준화된 가변전송률 음성 코덱이다. SMV 전송 환경과 상태에 따라서 Rate 1 (8.55 kbps), Rate 1/2 (4.0 kbps), Rate 1/4 (2.0 kbps)과 같은 전송률을 가진다. 또한 전송률과 음질 사이의 절충관계를 고려하여 4개의 동작모드를 갖는다. 따라서 SMV는 다양한 평균 전송율과 동작모드를 가지므로 CDMA 시스템의 성능과 음질간의 관계에서 효과적으로 성능을 조절 할 수 있다[5~7].

본 논문에서는 전송된 음성신호를 분석하여 이를 보이스 피싱 검출을 위한 효과적인 특징벡터를 선별하기 위해 SMV 복호화 과정에서 자동적으로 추출되는 특징들을 선별하였다. 선별된 특징벡터는 지난 보이스 피싱 검출에 관한 우리의 연구에서 사용되어진 특징벡터를 선택하였으며, 다음과 같다[3, 7].

1. 반사계수 (Reflection Coefficients, RC)

자기 상관함수 (R)와 LPC 계수 (α)를 이용하여 계산한다.

$$k_m(i) = -\frac{R_m(i) + \sum_{j=1}^{i-1} \alpha_j^{i-1} \cdot R_m(i-j)}{E^{i-1}} \quad (1)$$

2. LSF의 첫번째 계수 (First-LSF)

SMV 부호화 과정으로 부터 전송된 비트스트림으로부터 구한 LSF의 첫 번째 계수를 사용한다[7].

3. 피치 지연(Pitch Lag)

SMV 부호화 과정에서 개회로 피치 검출을 사용하여 구해진 값을 SMV 복호화 과정에서 전송받아서 적응 코드북을 통해 계산된다.

4. 수정된 피치 지연(Corrected Pitch Lag, CPL)

SMV 부호화 과정에서 전송된 피치 지연을 SMV 복호화 과정에서 프레임 타입, 불량 프레임 지시변수 그리고 반사계수를 이용하여 새롭게 구해진다.

Ⅲ. 보이스 피싱 검출을 위해 제안된 알고리즘

지금까지 연구되어진 보이스 피싱 검출은 단순한 결정식을 사용하거나 일반적인 방법의 패턴인식 방법을 사용하였다[8~9]. 하지만 본 논문에서는 추출되어진 특징벡터의 비교 및 분석을 통해 효과적인 특징벡터를 찾아내고, 일반적인 방법의 패턴인식이 아닌 MCE를 사용하여 인식에 사용되어지는 모델에 변별적 가중치를 적용한 보이스 피싱 검출 알고리즘을 제안한다.

제안된 보이스 피싱 검출에 기반이 되는 Gaussian Mixture Model (GMM)은 Expectation Maximization (EM) 알고리즘을 기반으로 주어진 데이터 집합에 대한 분포밀도를 복수개의 가우시안 확률밀도함소로 모델링하는 패턴인식의 방법 중 하나이다[10~11]. 우리가 지난 연구에서 다양한 분석을 통해 선택한 보이스 피싱을 위한 효과적인 특징벡터를 N개의 D차원 특징벡터 $X = \{x_1, x_2, \ldots, x_N\}, x_t \subseteq R^D$라고 하면, M개의 혼합성분 (Mixture Component)으로 구성되는 가우시안 확률밀도함수를 기반으로 하는 우도 (Likelihood)는 다음과 같이 계산된다.

$$p(\vec{x_t}|\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x_t})$$
$$b_i(\vec{x_t}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x_t} - \mu_i)^T (\Sigma_i)^{-1} (\vec{x_t} - \mu_i)\right\} \quad (2)$$
$$0 \le p_i \le 1, \ \sum_{i=1}^{M} p_i = 1$$

여기서 GMM을 위한 진실 (Truth) 모델 λ_T와 거짓

(Lie) 모델 λ_L는 다음과 같이 가우시안 혼합 성분 밀도의 가중치 (Mixture Weight : p_i), 평균 벡터 (Mean Vector : μ_i) 그리고 공분산 행렬 (Covariance Matrix : Σ_i)로 구성된다.

$$\lambda = \{p_i, \mu_i, \Sigma_i\},\ i = 1, ..., M \tag{3}$$

구성되어진 각각의 모델 파라미터 λ_T와 λ_L는 EM 알고리즘을 사용하여 $p(x|\lambda') \geq p(x|\lambda)$가 되는 새로운 모델 λ'를 정해진 문턱 값에 도달할 때까지 반복하여 알맞은 모델을 선별하게 된다. 이렇게 선별된 모델을 기반으로 다음과 같은 보이스 피싱을 위한 1차적인 결정식을 만들 수 있다.

$$\Lambda = \log \frac{p(\vec{x}|\lambda_T)}{p(\vec{x}|\lambda_L)} \underset{\text{Lie}}{\overset{\text{Truth}}{\gtrless}} \eta \tag{4}$$

여기서 η는 보이스 피싱 검출을 위한 문턱값이며, λ_T는 진실 모델 그리고 λ_L는 거짓 모델을 나타낸다.
제안된 방법은 보이스 피싱을 검출하는 결정식에 MCE 기법을 적용하여 각각의 모델별 혼합성분에 분별적 가중치를 적용한 최적의 감정 별 모델을 만드는 것이며, 제안된 최종 결정식은 다음과 같이 나타낼 수 있다.

$$\Lambda^{\omega} = \log \omega_i \frac{p_i^T b_i^T(\vec{x})}{p_i^L b_i^L(\vec{x})} \underset{\text{Lie}}{\overset{\text{Truth}}{\gtrless}} \eta \tag{5}$$

여기서 Λ^{ω}는 제안된 MCE 기법을 통해 구해진 혼합성분별 가중치가 적용된 최종 결정식을 나타낸다. 최종 결정식을 위한 최적을 가중치 ω_i를 구하기 위해 Generalized Probilistic Descent (GPD) 기법을 사용하게 된다. 이러한 기법을 기반으로 실제 훈련 데이터의 $\Lambda^{\omega}(t)$에 대한 분류 오류 $D(t)$를 정의할 수 있다.

$$D(\Lambda^{\omega}(t)) = \begin{cases} -g_T(\Lambda^{\omega}(t)) + g_L(\Lambda^{\omega}(t)), & \text{if current frame is Truth frame} \\ -g_L(\Lambda^{\omega}(t)) + g_T(\Lambda^{\omega}(t)), & \text{if current frame is Lie frame} \end{cases} \tag{6}$$

여기서 t는 프레임 인덱스를 나타내며, g_T와 g_L은 다음과 같이 입력 데이터를 진실과 거짓 프레임으로 분류하기 위한 함수이다.

$$g_T(\Lambda^{\omega}(t)) = \Lambda^{\omega}(t) - \theta \tag{7}$$
$$g_L(\Lambda^{\omega}(t)) = \theta - \Lambda^{\omega}(t)$$

분류 오류 함수는 음수 값을 가질 경우 올바른 분류로 판별하며 이를 기반으로 다음과 같이 손실함수 L을 정의할 수 있다.

$$L = \frac{1}{1 + \exp(-\beta D(\Lambda^{\omega}(t)))} \tag{8}$$

여기서 β는 sigmoid 함수의 기울기를 나타내며 구하고자 하는 최종 결정식을 위한 최적 가중치 ω_i는 Generalized Probabilistic Descent (GPD) 알고리즘에 기반하여 손실함수 L의 값이 최소가 될 때 구해지게 된다.

IV. 실험결과 분석 및 비교

본 논문에서 제안한 변별적 가중치 기반의 보이스 피싱 검출 알고리즘의 성능 평가를 위해서 실제 보이스 피싱 음성을 수집하였다. 이 음성은 우리의 지난 논문에서 사용하였던 데이터 파일과 같은 것으로[3], 보이스 피싱 모델을 만들기 위해 보이스 피싱 가해자의 목소리 남자 5명의 약 3분, 여자 3명의 약 1분 30초 분량의 목소리를 모았으며, 일반적인 남자와 여자의 전화통화 목소리 또한 같은 양의 데이터를 모았다. 또한 같은 양의 데이터가 테스트를 위해 사용되어 졌다. 수집되어진 일반적인 전화통화 목소리와 보이스 피싱 가해자 목소리 데이터는 모두 8 kHz로 샘플링 하였으며, 16 bit로 양자화 하였다.

제안된 보이스 피싱의 성능을 테스트하기 위해서 II 장에서 기술되어진 지난 우리의 연구에서 보인 효과적

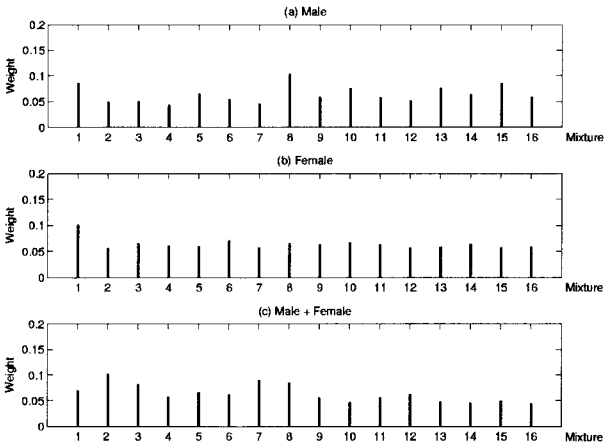

그림 1. GMM의 혼합 성분에 따른 가중치 분포
Fig. 1. Weights distribution according to Gaussian mixtures.

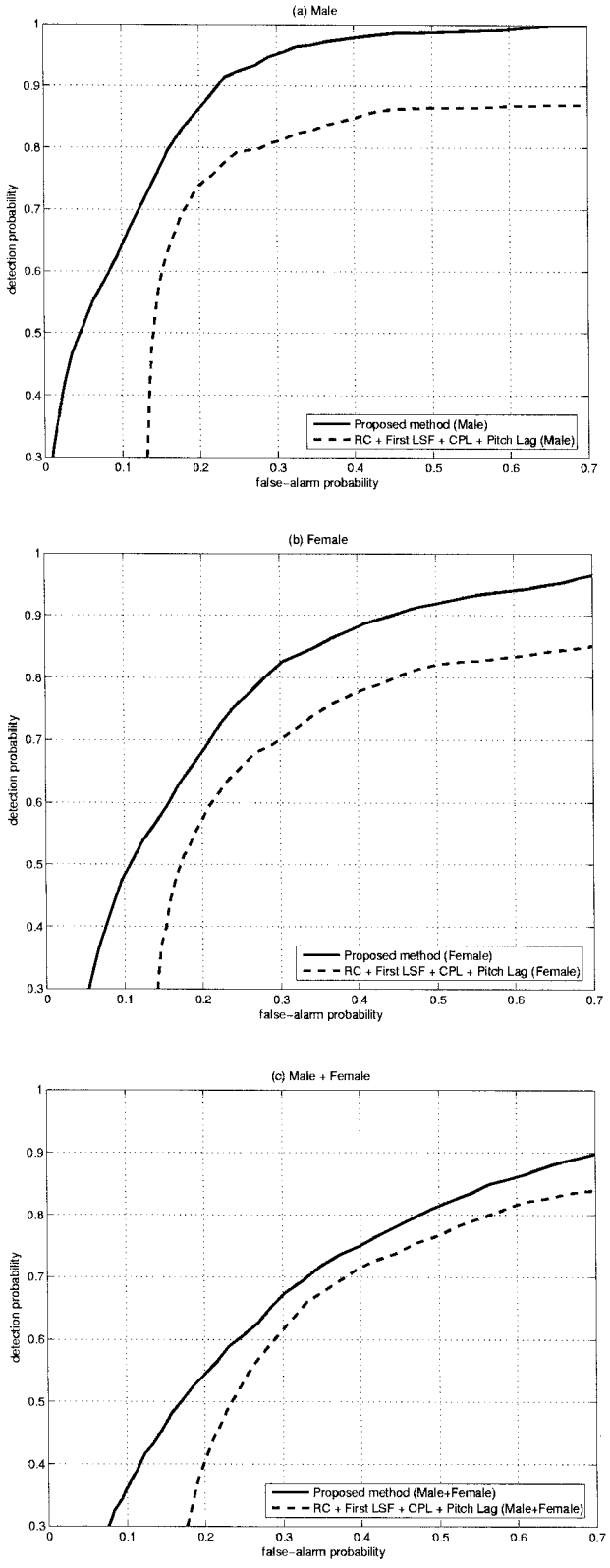

그림 2. ROC에 기반한 인식 성능 비교 (a) 남자 (b) 여자 (c) 남자 + 여자

Fig. 2. The voice phishing detection performance based on ROC. (a) Male (b) Female (c) Male + Female.

표 1. 제안된 방법과 기존의 방법의 EER 결과

Table 1. EER result according to voice pishing method.

특징벡터	EER		
	Male	Female	M+F
제안 방법	17.50%	24.38%	31.49%
기존 방법[3]	22.89%	29.95%	33.60%

인 특징벡터를 사용하였으며, 16개의 혼합성분을 사용하여 GMM 모델 구성하였다.

그림 1은 본 논문에서 제안한 MCE를 이용하여 구한 혼합성분 별 가중치를 보여준다. 그리고 그림 2와 표 1은 각각 구해진 혼합가중치를 적용하여 새롭게 구성 한 모델이 사용된 Receiver Operating Characteristic (ROC) 커브와 Equal Error Rate (EER)의 결과를 보여준다.

인식 결과 보이스 피싱 검출을 위해 제안 되어진 방법이 기존의 연구되어진 동일 특징벡터와 단순한 GMM을 사용한 인식 방법에[3] 비해 EER과 ROC 전반에 걸쳐 우수한 성능을 보인 것을 알 수 있었다. 이는 기존의 일반적인 GMM 방법에서는 각각의 혼합성분별로 로그 우도 차이를 MCE를 사용함으로써 보다 효과적인 혼합성분을 찾고 이를 바탕으로 변별적 가중치를 적용하였기 때문에 미세한 차이를 보이는 혼합성분의 로그 우도는 그 비중을 줄이고 많은 차이를 보이는 혼합성분의 로그 우도의 비중을 늘림으로써 더욱 향상된 인식 결과를 보인 것이라 생각된다.

V. 결 론

본 논문에서는 효과적인 전화사기 예방을 위해 최소 분류 오류 기법에 기반한 보이스 피싱 검출 알고리즘을 제안하였다. 기존의 연구에서 알아낸 효과적인 특징벡터에 일반적인 GMM 방식이 아닌 MCE를 적용하여 변별적인 가중치가 적용된 혼합성분을 이용한 GMM을 사용하여 인식을 수행하였다. 실험 결과 본 논문에서 제안한 방법이 기존의 방법에 비해 우수한 성능을 보인 것을 알 수 있었다. 또한 남자와 여자를 따로 테스트 할 경우 혼합해서 할 경우 보다 더 많은 성능의 향상을 비추어 볼 때 보이스 피싱 검출 입력 단에 우수한 성능의 성별인식기를 추가한다면 보다 효과적인 인식이 될 것이라 생각된다. 또한 이에 그치지 않고 보다 효과적인

특징벡터와 인식 방법에 대한 다양한 시도와 연구가 진행 되어야 할 것이다.

참 고 문 헌

[1] Furedly J. J., Davis C., and Gurevich M., "Differentiation of deception as a psychological process:A psychophysiological approach," *Psychophysiology*, vol. 25, no. 6, pp.683-688, 1988.

[2] Ekman P., Friesen W. V., and Scherer K., "Body movement and voice pitch in deceptive interaction," *Semiotica*, vol. 16, no. 1, pp. 23-27, 1976.

[3] 이계환, 장준혁, "3GPP2 SMV 기반의 보이스 피싱 검출 알고리즘," 전자공학회, 제 45권, SP 편 제 4호, pp. 92-99, 2008.

[4] Kang S. -I., Jo Q. -H., Chang J. -H., "Discriminative Weight Training for A Statistical Model-Based Voice Activity Detection," *IEEE Signal Processing Letters*, vol. 15, pp. 170-173, 2008.

[5] Greer S. C., and Dejaco A., "Standardization of the selectable mode vocoder," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 953-956, 2001.

[6] Yang G., Shlomot E. B., Thyssen J., Huan-yu S., and Murgia C., "The SMV algorithm selected by TIA and 3GPP2 for CDMA applications," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 709-712, 2001.

[7] 3GPP2 Spec., "Software distribution for selectable mode vocoder (SMV), service option 56, specification," *3GPP2-C. Roa30-0, v3.0*, 2005.

[8] Daniel N., Kjell E., and Kornel L., "Emotion Recognition in spontaneous speech using GMM," *INTERSPEECH*, pp. 809-812, 2006.

[9] Tsang-Long P., Yu-Te C., and Jun-Heng Y., "Emotion recognition from Mandarin speech signals," *International Symposium on Chinese Spoken Language Processing*, pp. 301-304, 2004.

[10] Bishop C. M, *Neural networks for pattern recognition*, Oxford University Press, UK, 1995.

[11] Duda R. O., Hart P. E., and Stork D. G., *Pattern classification*, John Wiley & Sons, INC., 2001.

저 자 소 개

이 계 환(학생회원)
2007년 인하대학교 전자전기 공학부 학사.
2007년~현재 인하대학교 전자공학과 석사과정.
<주관심분야 : 디지털신호처리>

장 준 혁(정회원)
1998년 경북대하교 전자공학과 학사.
2000년 서울대학교 전기공학부 석사.
2004년 서울대학교 전기컴퓨터 공학부 박사.
2000년~2005년 (주)넷더스 연구소장
2004년~2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
2005년 한국과학기술연구원(KIST) 연구원
2005년~현재 인하대학교 전자공학부 조교수
<주관심분야 : 음성 신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>