1999년도 대한전기학회 하계학술대회 논문집 1999. 7.19-21

벡터 양자화를 위한 학습 알고리즘을 이용한 음성 전송 기술에 관한 연구

홍강유, 박상희
연세대학교 전기공학과

A study on the competitive learning algorithm for robust vector qantization to transmit speech signal

Kangyou Hong, Sang-Hui Park
Dept. of Electrical Eng. Yonsei Univ.

Abstract - The efficient representation and encoding of signals with limited resources, e.g., finite storage capacity and restricted transmission bandwidth, is a fundamental problem in technical information processing systems. Typically under realistic circumstances, the encoding and communication of message has to deal with different sources of noise and disturbances. In this paper, I propose a unifying approach to data compression by robust vector quantization, which explicitly deals with channel noise, and random elimination of prototypes. The resulting algorithm is able to limit the detrimental effect of noise in a very general communication scenario. In this paper, based on the robust vector quantization I have an experiment about speech coding.

1. 서 론

벡터 양자화(vector quantization)는 데이터 압축과 신호 처리의 중요한 주제이다. 이 방법은 일정한 크기의 코드북(codebook)으로 정보를 인코딩 하는 문제를 다룬다. 많은 경우에 소스의 확률밀도함수는 미리 알려져 있지 않지만, 학습 데이터 벡터 $X=\{x_i\in R^d: 1\le i\le n\}$의 표본 집단은 알 수 있다. 적응력이 있는 코드북의 설계는 코드북 벡터 혹은 원형으로 이루어진 코드북 $Y=\{y_a\in R^d: 1\le a\le m\}$을 찾는 일과 유도된 기대되어지는 일그러짐이 주어진 미분 가능한 일그러짐의 측정 D에 대하여 일그러짐이 최소가 되는 인코딩 mapping $c: R^d\rightarrow\{1,\cdots,m\}$를 찾는 일을 다룬다. 이것은 종종 실험적인 위험이나 평균 일그러짐을 최소화하면서 달성된다.

$$H^{vq}(c, Y; X)=\frac{1}{n}\sum_{i=1}^{n} D(x_i, y_{c_i}),\quad c_i\equiv c(x_i) \tag{1}$$

위에서 언급된 식은 이론적으로 중요할 뿐만 아니라 음성이나 영상 압축과 같은 중요한 응용분야들을 가지고 있다. c(x)가 신호의 공간을 분할하므로, H^{vq}는 y_a가 집단 $C_a=\{x_i\in R^d: c_i=a\}$를 대표하는 값으로 집단화하는 문제에 대응된다. 가장 유명한 디자인 방법은 LBG 알고리즘이다.

On-line 벡터 양자화 방법과 매우 비슷한 competitive 학습 알고리즘은 신경망 회로(neural network)와 관련된 논문들에서 심도 있게 논의되어 왔다. 이러한 논문들에서 원형 Y는 입력 신호 공간에서의 수용체의 중심과 같은 신경(neuron)의 시냅스의 가중치에 대응된다. 이러한 알고리즘들의 일반적인 특징은 각 학습 단계에 선행되고, neuron이 새로운 자극 x에 가중치를 적용시킬 수 있도록 하는 competitive 단계에 있다. 가장 단순하고 가장 원형에 가까운 학습 규칙으로는 "winner-take-all"(WTA) 학습이 있다. 이 학습방법에서 적응은 x에 가장 잘 어울리는 한 개의 neuron에 제한된다. WTA학습은 입력 공간의 다른 부분들을 각기 다른 neuron에 대응시킨다. WTA의 한계점들을 약화시킴으로서, soft competition은 하나 이상의 신경이 pattern x_{n+1}을 표현하도록 하는 것이다. 각각의 신경은 각각 개인적인 상관 적응 세기 $\rho_a^{(n+1)}\equiv\rho_a(x_{n+1})$을 가지고 있다. 이 논문에서 soft competitive 학습 규칙의 일반적인 형태는 아래의 식 (1.2) 와 같이 표현될 수 있다.

$$Y_a^{(n+1)}=Y_a^{(n)}-\frac{1}{2}\rho_a^{(n+1)}\frac{\partial}{\partial y_a^{(n)}}D(x_{n+1}, Y_a^{(n)}) \tag{2}$$

이 논문에서는 먼저 아래의 두 가지 중요 신경망 회로 알고리즘이 다른 환경의 잡음(noise)과 대역폭의 변화하는 조건에서 robust 인코딩(encoding) 하는 툴이 유도되는 것을 보여줄 것이다.

1) the self-organizing feature map(SOFM)
2) the "neural-gas"(NG) 알고리즘

2. 잡음이 없는 통신 채널에서의 벡터 양자화.

2.1 잡음이 없는 채널을 위한 코드북 설계

x는 $y_{c(x)}$로 인코드되고, 지표 c(x)는 채널을 통하여 전송되고, 이것은 손상 없이 얻어져서 $y_{c(x)}$가 저장되는 경우이다. 이러한 통신 개념은 식 (1)에 나온 목적함수 H^{vq}에 의하여 표현될 수 있다. LBG 알고리즘은 인코딩 함수 c와 코드북 Y에 대하여 H^{vq}의 최소 값을 변경한다. 그 결과 아래와 같은 식이 나오게 된다.

$$c_i=\arg\min D(x_i, y_a),\ \sum_{i=1}^{n}\delta(a, c_i)\frac{\partial D(x_i, y_a)}{\partial y_a}=0 \tag{3}$$

여기서 δ는 Kronecker's delta 함수를 나타낸다.

2.2 잡음이 없는 채널을 위한 On-Line 학습

LBG나 다른 일련의 학습 알고리즘에 의한 H^{vq}와 같은 경험에 의한 위험 함수의 최소화는 모든 학습 데이터 X가 최적화 과정 이전에 주어져야 할 필요가 있다. On-line 학습 규칙을 얻기 위해서, n 번째와 n+1 번째의 데이터에 대한 중심 조건의 차를 구한다.

$$Y_a^{(n+1)}-Y_a^{(n)}=\frac{\delta(a, c_{n+1}^{(n+1)})}{n_a^{(n+1)}}x_{n+1}+\sum_{i=1}^{n}\hat{\omega}_{ia}x_i \tag{4}$$

그리고,

$$\hat{\omega}_{ia}\equiv\frac{\delta(a, c_i^{(n+1)})n_a^{(n)}-\delta(a, c_i^{(n)})n_a^{(n+1)}}{n_a^{(n)}n_a^{(n+1)}} \tag{5}$$

여기서, $n_a^{(n)}\equiv\sum_{i=1}^{n}\delta[a, c_i^{(n)}]$는 y_a에 대한 인코딩 주파수이고, 위 첨자는 n 번째와 n+1 번째 데이터의 인코딩 함수와 코드북 벡터를 각각 나타낸다. On-line 설정에서는 과거의 데이터 값의 인코딩을 다시 계산하는 것은 복잡하므로, $c_i^{(n+1)}\approx c_i^{(n)}$로 근사화하고, 모든 불필요한 위첨자들을 제거한 뒤, 아래와 같은 on-line 식에 도달할 수 있다.

$$y_\alpha^{(n+1)} = y_\alpha(n) + \rho_\alpha^{(n+1)}(x_{n+1} - y_\alpha^{(n)}), \rho_\alpha^{(n)} \equiv \frac{\delta(\alpha, c_n)}{n_\alpha^{(n)}} \quad (6)$$

3. 제한이 있는 통신 채널에서의 벡터 양자화.

3.1 잡음이 있는 통신 채널에서의 벡터 양자화.

3.1.1 Source-Channel 코딩

Source-channel 코딩이라고 알려진 벡터 양자화 문제의 중요한 확장은 코드북을 설계하는 단계에서 잡음이 있는 통신 채널을 고려하는 것이다. 알려진 잡음 특성을 가진 불연속적이고 메모리가 없는 채널의 경우에 한정하였다. 전송 행렬 S의 원소인 $s_{\nu|\alpha}$를 채널을 통하여 α를 보낸 다음에 지표 ν를 받을 가능성으로 설정한다. 목적 함수 H^{svq}는 다음과 같이 일반화된다.

$$H^{svq}(c, Y; X) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\nu=1}^{m} s_{\nu|c_i} D(x_i, y_\nu) \quad (7)$$

만약 $s_{\nu|\alpha}$가 높다면, 손상된 지수로부터 재배치하는데 따르는 해로운 효과를 제한하기 위하여 y_α와 y_ν를 서로 가까운 곳에 놓는 것이 유리하다. 이것은 다음과 같은 안정된 식으로 직접적으로 표현될 수 있다.

$$c_i = \arg\min D_{i\alpha}^S, \quad \sum_{i=1}^{n} s_{\nu|c_i}\frac{\partial D(x_i, y_\nu)}{\partial y_\nu} = 0 \quad (8)$$

여기서 $D_{i\alpha}^S \equiv \sum_{\nu=1}^{m} s_{\nu|\alpha} D(x_i, y_\nu)$이다.

3.1.2 일반화된 SOFM 학습 규칙

SOFM에 대한 관계는 H^{svq}에 대응하는 on-line 학습 수식을 유도해 냄으로서 보다 명확하게 될 수 있다. 잡음이 없는 경우와 마찬가지로 식들을 전개하여, $\rho_\alpha^{(n)} \equiv s_{\alpha|c_n}/n_\alpha^{(n)}$ 으로 설정하여 식 (6)의 일반화된 식을 얻었다. 여기서 c_n은 식 (3)에서의 WTA 규칙들 대신에 주어졌고, $n_\alpha^{(n)} \equiv \sum_{i=1}^{n} s_{\alpha|c_i}$이다. 위의 soft competitive 학습 규칙들은 최소한의 기대 오차를 가진 신경으로 "winner"를 정의함으로서 기본적인 SOFM 학습을 수정하였다.

3. 2 Robust 벡터 양자화

3.2.1 원형 제거 모델

Robust 벡터 양자화이라 불리는 기본적인 벡터 양자화 모형의 또 다른 중요한 확장은 코드북 벡터의 불규칙한 삭제를 다룬다. 채널 잡음을 포함하는 robust 벡터 양자화의 인코딩/전송 방법은 그림 1 에 나타나 있다.

이 통신 모형에서, 코드북 설계는 어떤 원형들은 일시적인 코드북의 축소 $Y^{(t)} \subseteq Y$ 때문에 인코딩 시간 t에 사용할 수 없을지도 모른다는 문제점들을 다룬다.

3.2.2 일반화된 "Neural-Gas" 모형

불규칙적인 제거는 사용 가능한 code alphabet에 대하여 불확실한 것을 인코딩 하게 만든다. 이것은 기본적으로 잡음이 있는 전송 채널 때문에 일어나는 불안정을 디코딩 하는 것과는 기본적으로 다르다. 모든 데이터 벡터에 있어서, {1,…,m} 에 전단사로 대응되는 순위 r_i 를 도입하였다. 예를 들면, x_i를 인코딩 하기 위해서 u번째를 선택하고, 반면에 역함수 $\bar{r}_i(\alpha) \equiv r_i^{-1}(\alpha)$ 은 y_α의 순위를 나타내는, $r_i(u)$는 순위 u를 가진 원형지표를 나타낸다. 시간 t에서 데이터를 인코딩 하기 위하여, 코드북의 사용 가능한 $Y^{(t)} \subseteq Y$ 인 부분에서 가장 낮은 순위

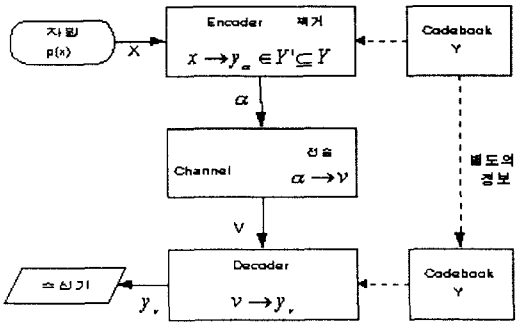

그림 1. 채널 잡음과 코드북 벡터의 불규칙적인 제거를 갖는 벡터 양자화.

r을 갖는 y_α는 x_i를 인코드 하기 위하여 사용된다 ; 더욱 공식적으로 표현하자면, $c_i \equiv \arg\min_{\{\alpha | y_\alpha \in Y^{(t)}\}} \bar{r}_i(\nu)$ 이다.

y_α에 대한 독립된 제거 잡음 ε_α의 경우를 생각해보자. 즉, 원형 y_α는 인코딩 시간 t에 사용 가능할 확률이 $(1-\varepsilon_\alpha)$를 가지며 원형들의 집합에 대한 공동의 가능성은 인수분해 된다. 이러한 경우에, robust 벡터 양자화에 대한 목적함수는 아래와 같이 주어진다.

$$H^{rvq}(r, Y; X) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\alpha=1}^{m} p_{\alpha|r_i} D(x_i, y_\alpha) \quad (9)$$

그리고,

$$p_{\alpha|r_i} \equiv \frac{1-\varepsilon_\alpha}{1-\prod_{\nu=1}^{m}\varepsilon_\nu}\prod_{u=1}^{\bar{r}_i(\alpha)-1}\varepsilon_{r_i(u)} \quad (10)$$

여기서 $p_{\alpha|r_i}$는 주어진 순위 r_i에 대한 y_α의 인코딩 확률로 해석될 수 있다. 일정한 제거 확률 $\varepsilon_\alpha \equiv \varepsilon$의 경우, 목적함수는 neural-gas 모형 H^{ng}로 단순하게된다. 여기서 $p_{\alpha|r_i} \propto \varepsilon^{\bar{r}_i(\alpha)}$ 이다. H^{ng}에 대하여, $D(x_i, y_\alpha)$의 기여에 대한 중심은 단지 y_α의 순위에만 의존한다. H^{rvq}의 안정화 조건은 아래와 같다.

$$D(x_i, y_{r_i(u)}) \le D(x_i, y_{r_i(v)}), \text{ 단 } u < v \quad (11)$$

그리고,

$$\sum_{i=1}^{n} p_{\alpha|r_i}\frac{\partial D(x_i, y_\alpha)}{\partial y_\alpha} = 0 \quad (12)$$

여기서 적용 능력은 $\rho_\alpha^{(n)} = p_{\alpha|r_n} / \sum_{i=1}^{n} p_{\alpha|r_i}$이다.

4. Robust 벡터 양자화 모형의 통합

여태까지는 기본적인 코드북을 설계하는 문제의 두 가지 확장을 각각 소개하였다. 이제 어떻게 이러한 다른 모델들이 합하여 질 수 있는지 알아보겠다.

정하여진 통신 행렬 S에 대하여 잡음이 있는 통신 모형과 원형 제거의 혼합은 아래와 같이 유도되어질 수 있다. 식 (4.3)과 식 (4.4)의 안정화 조건은 다음과 같이 수정될 수 있다.

$$D_{ir_i(u)}^S \le D_{ir_i(v)}^S, \sum_{i=1}^{n} s_{\nu|c_i} p_{\alpha|r_i}\frac{\partial D(x_i, y_\nu)}{\partial y_\nu} = 0 \quad (13)$$

비슷하게 on-line 최적화에 대한 적용 중심도 SOFM과 NG 알고리즘을 혼합하여 얻어진다.

아래와 같은 과정을 실행하면 제안된 알고리즘의 해가 나온다.

과정 1. 모든 y_α를 불규칙하게 초기화한다.

과정 2. i=1,…,n 일 때, 과정 2-1로

과정 2-1. α=1,…,m ; ν=1,…,m 일 때 ; 단 $\nu \neq \alpha$

중심 벡터 w_{av}^{i}를 구한다

과정 3. a =1,…,m 일 때

중심 규칙 식 (13)에 따라서 y_a 를 갱신한다.

과정 4. 위의 과정을 수렴할 때 까지 반복한다.

과정 5. 모든 y_a에 작은 크기를 갖는 가우시안 잡음 (Gaussian noise)을 첨가한다.

5. 실험 및 결과 고찰

Fourier 변환된 음성 신호를 가지고 앞에서 설명된 on-line 경우에서의 실험을 하였다.

실험에 사용된 데이터는 16KHz로 표본화된 '바람과 햇님'의 이야기 대본이다. 16개의 표본마다 프레임을 나누어 처음 1024개의 프레임으로 1024개의 크기를 갖는 코드북을 작성하였다. 이 데이터를 Fourier 변환을 하여 노이즈를 첨가하고 제안한 알고리즘으로 인코딩 후 복원하는 방법으로 실험하였다. 이 실험에서 robust 벡터 양자화로 설계된 코드북이 LBG 알고리즘과 비교할 때 학습 데이터에서 뿐만이 아니라 실험 데이터에서도 뛰어난 성능을 보여준다. 모든 코드북 학습 데이터와 실험 데이터 사이에 주목할만한 약 2 dB의 SNR 차이를 보여주었다. 이것은 데이터의 적응 때문에 나타난다. 그림 2는 같은 데이터를 가지고 robust 벡터 양자화 설계의 장점을 보여주고 있다. 코드북의 제거 확률은 $\varepsilon_a = \varepsilon = 0.3$으로 설정되었다. 채널 잡음은 독립되게 1%의 작은 error 비율을 주었다. 전송 error 확률은 약 8% 정도였다. LBG 알고리즘으로 만든 코드북에 비하여 SNR 곡선이 학습 데이터와 실험 데이터 모두에서 약 3dB의 성능 향상을 보였다. 이것은 잡음에 의한 SNR 손실의 50% 정도가 robust 설계 과정에 의하여 보상되어진 것을 의미한다. 성능의 향상은 예로 들어놓은 음성 데이터에서 확인할 수 있다.

표 1 채널 잡음과 제거 잡음이 모두 있을 경우의 평균 SNR

	잡음이 없는 경우	WTA	robust 벡터 양자화
평균 SNR	37.07	31.39	34.32

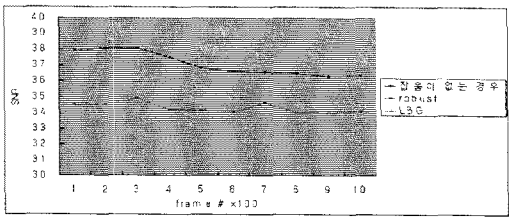

그림 2 음성 신호에 대한 robust 벡터 양자화의 성능 비교

6. 결 론

본 논문에서는 데이터를 인코딩하고 전송하는데 있어서 두 가지 다른 형태의 한계에 robust 벡터 양자화 모형을 제안하였다. 첫 번째는 잡음이 있는 전송 채널이고 두 번째는 코드북 벡터의 제한된 사용 가능성이다. 각각의 제한들은 다음과 같은 알려진 competitive 학습 알고리즘과 관련이 있다. 첫 번째는 self-organizing feature map이고 두 번째는 "neural-gas" 알고리즘의 일반화된 형태이다.

통합된 모델은 동시에 다른 제한들을 다루는데 적합하다. 그러므로, robust 코드북 설계는 많은 실제적인 상황에서 적용될 수 있다. On-line 적응은 앞에서 언급된 neural net 갱신 규칙들에 의하여 실현될 수 있다. 현재 있는 일반적인 방법들과 본 논문에서 제안된 방법들을 비교해 볼 때 계산상의 복잡성은 미미한 반면, 복원시 손실은 훨씬 적은 것으로 나타났다.

본 논문에서 얻어진 결과는 다음과 같다. 다음과 같다. 첫째, Robust 벡터 양자화를 사용하여 코드북을 설계하였을 때와 LBG 알고리즘을 사용하였을 때와 비교하여 보면 채널 잡음만이 있는 경우 WTA 알고리즘과 비교하여 볼 때, 1 dB, 제거 잡음만이 있는 경우에는 0.8 dB ~ 2.0 dB 정도의 성능 향상, 그리고 채널 잡음과 제거 잡음이 모두 있는 경우에는 일반적인 LBG 알고리즘과 비교할 때 약 3dB 정도의 성능 향상이 있었다. 둘째, 일반적이 WTA 학습은 noise 가 증가할 경우 성능의 변화가 robust 벡터 양자화 알고리즘을 사용한 경우보다 심하다.

[참 고 문 헌]

[1] R. M. Gray, "Vector quantization," IEEE Acoust, Speech, Signal Processing mag., pp. 4-29, Apr. 1984.

[2] Thomas Hofmann and Joachim M. Buhmann, "Competitive learning algorithms for robust vector quantization" IEEE Transactions on signal processing. pp 1665-1675 June 1998.

[3] T. Lookabaugh, E. A. Riskin, P. A. Chou, and R. M. Gray, "Variable rate vector quantization for speech, image, and video compression," IEEE Trans. Commun., vol. 41, pp. 186-199, 1993.

[4] Y. Linde, A Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol. COMM-28, pp. 84-95, 1980.

[5] T. Martinetz, S. G. Berkovich, and K. J. Schulten, "Neural-gas network for vector quantization and its application to time-series prediction," IEEE Trans. Neural Networks, vol. 4, pp. 558-569, 1993.

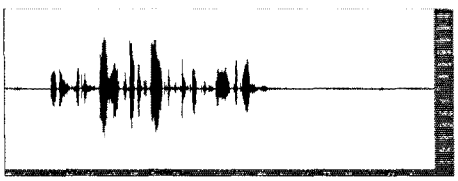

그림3 실험에 사용된 음성신호

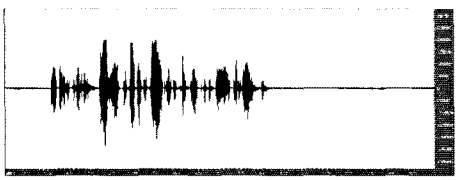

그림 4 채널 잡음과 제거 잡음이 있을 때 LBG 알고리즘을 사용하여 복원한 음성

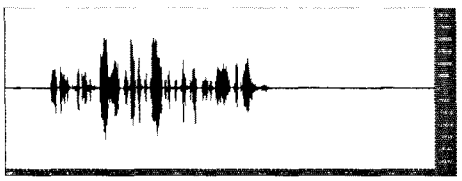

그림 5 채널 잡음과 제거 잡음이 있을 때 robust VQ 알고리즘을 사용하여 복원한 음성