

Logistic Regression 방법을 이용한 천이 신호 식별 알고리즘 및 성능 분석

*허순철⁰, *김진영¹, *윤병수², *남상원³, **오원진⁴
*한양대학교 전기공학과, **국방과학연구소

On the Performance Analysis of a Logistic regression based transient signal classifier

*Sun-Cheol Heo, *Jin-Young Kim, *Byoung-Soo Yoon, *Sang-won Nam, **Won-Cheon Oh
*Dept. of Electrical Engineering, Hanyang Univ., **Agency for Defence Development

Abstract - In this paper, a transient signal classification system using logistic regression and neural networks is presented, where four neural networks such as MLP, MLP-Class, RBF and LVQ are utilized to classify given transient signals, based on the logistic regression method. Also, some test results with experimental transient signal data are provided.

I. 서론

최근 음성 인식, 문자 인식, 군사 관련 분야, 그리고 일반적인 설비나 정밀 기기의 오류 검출과 같은 상업분야 등에서 신호를 자동으로 분류할 수 있는 식별기(즉, 식별 알고리즘)에 대한 관심이 높아지고 있다[1, 2]. 특히, 전통적인 패턴 분류 방식은 한 종류의 특징 벡터와 특정 식별 알고리즘을 사용하여 주어진 패턴의 종류를 결정하였다. 그러나, 많은 종류의 패턴을 분류하여야 하고, 잡음에 의해 왜곡된 신호를 식별할 경우 하나의 식별기로 높은 신뢰성을 갖는 식별 결과를 얻기는 어렵다. 따라서, 서로 다른 형태의 식별기를 함께 사용하여 높은 식별률과 신뢰성을 갖도록 하는 식별기 결합에 대한 연구가 최근 활발히 진행되고 있다[4, 5, 6].

식별기 결합 알고리즘은 개별 식별기들의 상호 보완 작용에 의해, 각 식별기의 장점을 이용하고, 단점을 서로 보완함으로써 인식률 향상과 높은 신뢰성을 가져야 진정한 결합 알고리즘이라 할 수 있다.

본 논문에서는 logistic regression에 근거한 식별기 결합 알고리즘을 이용하여, 여러 식별기들을 각 출력 scale에 상관 없이 결합할 수 있고, 보다 높은 신뢰성을 갖는 일반적인 식별 시스템의 구성을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 기본 식별 알고리즘과 식별기 결합 방법에 대해 약술하고, 3절에서는 실험적으로 측정된 천이 신호를 적용하여 제안된 식별 시스템의 성능을 테스트하였으며, 끝으로, 4절에 결론을 보인다.

2. 식별 시스템 구성

신호를 자동으로 분류하기 위한 식별 시스템은 그림 1과 같이 구성하였다. 신호가 식별기에 들어오면 신호의 특징 벡터가 각 식별 알고리즘의 입력으로 들어가게 되고, 각 식별 알고리즘은 독립적으로 신호를 분류한다. 각 식별 알고리즘으로 부터

나온 결과를 식별기 결합 알고리즘에 따라 결합하여 최종적인 판단을 하여 신호를 분류하게 된다[4, 5, 6].

2.1 기본 식별 알고리즘

본 논문에서는 신호에 대한 수학적인 사전 지식을 요구하지 않고 일반화하는 능력이 뛰어난 MLP(Multilayer Perceptron), RBF(Radial Basis Function), MLP-Class, LVQ(learning Vector Quantization) 등 네 종류의 신경 회로망을 기본 식별 알고리즘으로 선정하였다.

- MLP[3, 7]는 순방향 지도학습 신경 회로망으로써 여러 역 전파법으로 가중치를 수정하는 신경 회로망이다. 본 논문에서는 입력층, 은닉층, 출력층의 3개 층으로 MLP을 구성하였으며, 은닉층과 출력층 노드(node)에서 비선형 함수(sigmoid function)을 통과하게 하여 [0, 1] 사이의 값을 출력한다.

- RBF[7]도 입력층, 은닉층, 출력층의 3개 층으로 구성하였고, 은닉층에 미리 중심과 폭을 고정시킨 방사상 대칭인 함수를 사용하였으며, 최종 출력층에서 비선형 함수(sigmoid function)를 통과하게 하여 [0, 1] 사이의 출력을 얻는다.

- MLP-Class는 분류하고자 하는 각 부류에 대응하는 하나씩의 MLP를 모델화(그림 2 참조)한 신경 회로망이다.

- LVQ[8, 9]는 벡터 양자화의 일종으로, 몇몇의 codebook 벡터 또는 참조 벡터들로 패턴 공간을 분할하여 대표하게 한 알고리즘이다.

특히, MLP, RBF, MLP-Class는 [0, 1] 사이의 동일한 scale의 출력을 내며, LVQ는 입력 패턴과 참조 벡터들과의 거리를 출력으로 나타낸다.

2.2 식별기 결합 알고리즘

지금까지 발표된 식별기 결합 방법에는 Entropy-Based Integrator(ENT), Heuristic Combination of Confidence Factor(CF), Averaging(AVE), Majority Vote(MVOTE) 등이 있다[4, 6]. 이 중 AVE, ENT, CF 등의 결합 방법은 서로 같은 scale의 출력(즉, [0, 1] 사이의 아날로그 출력)을 보이는 식별기들의 결합에는 적합하지만, 서로 다른 scale의 출력(즉, 입력 패턴과 참조 벡터 사이의 거리 값, 확률 값 등)을 나타내는 식별기들을 결합할 경우, 각 식별기 출력 scale의 불일치 때문에 이 결합 방법들의 사용에 어려움이 있다. 반면, MVOTE 방법은 서로 다른 scale의 출력을 내는 식별기들의 결합에 사용할 수 있으나, 각 식별기의 식별 성능에 관계없이 개개 식별기를 똑 같은 비중으로 다루고 있다. 위에서 언급된 결

한 방법 중 대표적인 ENT에 대해 약술하고, 새로운 결합 방법인 Logistic Regression 방법 [5]에 대해 알아본다.

2.2.1 Entropy-Based Integrator (ENT)

ENT 방법은 각 출력 노드들의 값의 차가 큰 식별기, 즉 분류가 잘 된 식별기에는 큰 가중치를 주고, 각 출력 노드들의 값들이 비슷한 식별기에는 적은 가중치를 주어 신호를 분류하는 방법이며 다음 식과 같이 표현된다[4].

$$y_{ci}^* = \frac{y_{ci}}{\sum_c y_{ci}} \quad (1)$$

$$H(c) = \frac{1}{n} \sum_{i=1}^n \frac{y_{ci}}{\sum_c y_{ci}^* \ln y_{ci}^*} \quad (2)$$

여기서, c 는 폐턴 부류(class)를, i 는 각 식별기를, n 은 전체 식별기 개수를 나타내는 침자이고, y_{ci} 는 폐턴 부류 c 에 대한 식별기 i 의 출력을, y_{ci}^* 는 식별기 i 에 있어서 폐턴 부류 c 의 출력 y_{ci} 를 비례 축소화(scaling)한 값을 나타낸다. 가장 큰 $H(c)$ 값을 갖는 부류를 입력 신호의 부류로 할당한다.

2.2.2 Logistic Regression (LR)

Logistic Regression 방법[5]은 서로 다른 scale의 출력을 내는 식별기들을 체계적으로 결합하기 위해, 입력 폐턴에 대한 각 식별기의 출력을, 분류하고자 하는 각 부류에 대한 순위로 나타내어 최종 판단을 위한 입력으로 이용한다. 특히, 식별기 상호간의 식별 성능을 고려하여 각 식별기의 식별 능력에 따른 가중치를 부여함으로써 좀 더 효율적인 식별이 가능하도록 한 결합 방법으로 그 내용은 다음과 같다.

하나의 입력 폐턴이 주어졌을 때 그 입력 폐턴이 속하는 부류 C 에서 값 1을 갖고, 그렇지 않은 부류에서 값 0을 갖는 이진 변수 Y_e 를 설정한다. 식별기 C_1, C_2, \dots, C_m 에 의해 어떤 부류에 할당된 순위가 $X = (x_1, x_2, \dots, x_m)$ 일 때 Y_e 가 1이 되기 위한 확률은 $P(Y_e = 1|X) = \pi(X)$ 이다(식(3) 참조) [5]. 이 때 식별기 C_i 에 의해 어떤 부류가 최고 순위로 평가되면 x_i 는 최대값을 갖는다. 결국, 각 식별기의 부류별 순위는 분류하고자 하는 각 부류의 Y_e 값을 알아내는데 사용되는 임의 변수(random variables)라 할 수 있고, 이 순위들이 Y_e 값에 미치는 영향은 regression function을 이용하여 모델화할 수 있고, 다음의 logistic function $\pi(X)$ 로 나타낼 수 있다.

$$\pi(X) = \frac{\exp(a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m)}{1 + \exp(a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m)} \quad (3)$$

$$\text{즉, } L(X) = \log \frac{\pi(X)}{1 - \pi(X)} = (a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m)$$

여기서, $L(X)$ 는 log-odd, 또는 logit이라 하며, X 와 선형 관계에 있고, $a, b = (b_1, b_2, \dots, b_m)$ 은 상수 파라미터이며, least-squares 방법으로 이 모델 파라미터들을 추정할 수 있다[5]. 파라미터들의 상대적 크기는 logit 값에 미치는 각 식별기들의 상대적 중요도를 나타낸다. 즉, 이 파라미터들은 각 식별기들의 순위값에 대한 가중치로 사용되며, 가장 큰 logit 값을 갖는 부류를 그 입력 폐턴의 부류로 판단한다.

3. 실험 및 결과

그림 3은 실험에 사용된 8 부류의 천이 수중 소음 신호를 보여주고 있다. 본 실험에서는 각각의 신호에 0dB, 5dB, 10dB, 20dB(SNR 값)의 잡음을 섞어 원 신호를 포함하여 각 천이 신호별로 45개씩의 실험 데이터를 얻었으며, 20개 폐턴(SNR당 5개씩)을 기본 식별 알고리즘의 학습에, 나머지 25개 폐턴(SNR당 6개씩 및 원 데이터 포함)을 테스트에 사용하였다. 또한, 신호를 20개의 시간 간격(time slice)으로 나누고 각 slice당 11개의 AR계수를 구하여 총 220개의 AR계수를 입력 폐턴(특징 벡터)으로 사용하였다.

첫번째 실험(실험 I)은 그림 3의 (a) can (b) cow fish (c) pilot whale (d) balloon (e) seismic 등 5 부류의 천이 신호 식별에 관한 것으로, 전체 폐턴 중 100개 폐턴을 학습에 나머지 125개 폐턴을 테스트에 사용하였다. 두번째 실험(실험 II)은 실험 I의 5 부류 천이 신호에 (f) sperm whale (g) load fish (h) trigger fish의 3 부류를 더하여, 전체 폐턴 중 160개 폐턴을 학습에 200개 폐턴을 테스트에 사용하여 실시하였다.

실험 결과 분류한 부류가 적어 구별이 쉬운 경우(실험 I) 개별 식별기와 식별기 결합 알고리즘 모두 비슷한 높은 인식률(97.6%-99.8%)을 보이고 있으나(표 1 참조), 폐턴 분포가 복잡해질 경우(즉, 분류한 부류가 증가된 경우: 실험 II 및 표 2 참조) 개별 식별기의 인식 결과(93.5%-96.5% 인식률)보다 여러 종류의 식별기 결합 식별 알고리즘이 더 우수한 인식률(97-98% 인식률)을 나타내었다. 특히, 식별기 출력 scale에 상관없이 높은 인식률을 갖는 식별기들을 결합할 수 있는 logistic regression 방법(표 2: LR 1, LR 2, LR 3: 98% 인식률)이, 같은 출력 scale을 갖는 식별기들을 결합할 수 있는 ENT(97% 인식률)보다 더 유용함을 알 수 있다.

4. 결론

본 논문은 신호의 특징 벡터를 입력으로 하여 신호를 자동으로 식별할 수 있는 식별기 구성을 중점을 두었다. 특히, 기본 식별 알고리즘으로 MLP, RBF, LVQ, MLP-Class와 같은 신경 회로망을 선정하였으며, 식별기 결합 알고리즘으로 ENT와 Logistic Regression 방법을 기술하였다. 최종 식별 시스템은 기본 식별 알고리즘을 병렬로 연결하여 각 알고리즘 출력 결과를 logistic regression 결합 알고리즘이 따라 결합하였다. 또한, 특정 천이 신호를 이용한 실험에서, 본 논문에서 제시된 식별기 결합 방법이 개별 식별기 보다 우수함을 입증하였고, 식별기 출력 scale에 상관없이 식별기들을 결합할 수 있는 logistic regression 방법의 유용성을 확인하였다. 앞으로 보다 우수한 식별 시스템을 구성하기 위해서는 식별 시스템 구성의 개선 뿐 아니라 입력 신호를 가장 잘 대표할 수 있는 특징 벡터 추출에 대한 연구 또한 요구된다.

5. 참고문헌

- [1] J. Ghosh, L. Deuser, and S. Beck, 'A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals,' IEEE J. Ocean Eng., vol. 17, no. 4, pp. 351-363, 1992.
- [2] A. Kundu, G. C. Chen, and C. E. Persons, 'Transient

- sonar signal classification using hidden Markov models and neural nets,' IEEE J. Ocean Eng., vol. 19, no. 1, pp. 87-99, Jan. 1994.
- [3] J. A. Freeman and D. M. Skpura, *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison Wesley, 1991.
- [4] J. Ghosh, S. Beck, and C. C. Chu, 'Evidence combination techniques for robust classification of short-duration oceanic signals,' SPIE Conference on adaptive and Learning Systems, Orlando, FL, Apr. 1992
- [5] T. K. Ho, J. J. Hull, and S. N. Srihari, 'Decision Combination in Multiple Classifier Systems,' IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 1, pp. 66-75, Jan. 1994.
- [6] 윤병수, 양성칠, 남상원, 오원천, '신경회로망을 이용한 신호 자동식별기 구현 및 성능 분석,' 대한전기학회 추계학술대회 논문집, pp. 397-399, 1994.
- [7] P. D. Wasserman, *Advanced Methods in neural Computing*: Van Nostrand Reinhold, New York, 1993.
- [8] T. Kohonen, 'The self-organizing map,' Proc. of the IEEE, vol. 78, no. 9, pp. 1464-1480, 1990.
- [9] 양성칠, 남상원, 오원천, 'LVQ와 결정 트리를 결합한 패턴 분류 알고리즘,' 대한전기학회 하계학술대회 논문집 B, pp. 1037-1039, 1994.

그림 1. 5 부류 천이 신호 식별 결과(실험 I)

식별 시스템		classification rate
개별 식별기	MLP	123/125 (98.4%)
	MLP-Class	123/125 (98.4%)
	RBF	122/125 (97.6%)
	LVQ	123/125 (98.4%)
식별기 결합	BNT	123/125 (98.4%)
	LR 1	123/125 (98.4%)
	LR 2	123/125 (98.4%)
	LR 3	124/125 (99.8%)

그림 2. 8 부류 천이 신호 식별 결과(실험 II)

식별 시스템		classification rate
개별 식별기	MLP	190/200 (95.0%)
	MLP-Class	193/200 (96.5%)
	RBF	187/200 (93.5%)
	LVQ	192/200 (96.0%)
식별기 결합	BNT	194/200 (97.0%)
	LR 1	196/200 (98.0%)
	LR 2	196/200 (98.0%)
	LR 3	196/200 (98.0%)

* BNT : MLP, MLP-Class, RBF 결합

* LR 1: MLP, MLP-Class, RBF 결합

* LR 2: MLP, MLP-Class, LVQ 결합

* LR 3: MLP, MLP-Class, RBF, LVQ 결합

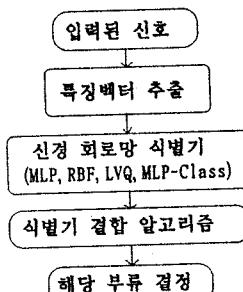


그림 1. 식별 시스템

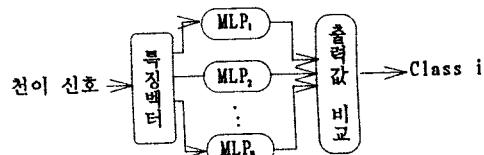


그림 2. MLP-Class

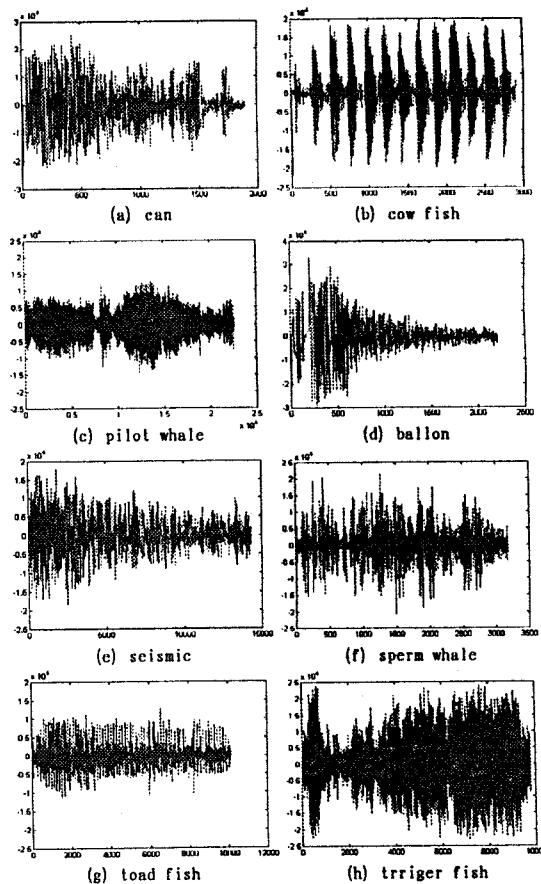


그림 3. 실험에 사용된 8 부류의 천이 수중 소음 신호