

PCA를 이용한 전력품질 특징분석

이진목, 홍덕표, 김수철, 최재호, 홍현문*
 충북대학교 전기전자 및 컴퓨터공학부, *동해대학교 전기전자공학과

A Feature Analysis of the Power Quality Problem by PCA

Jinmok Lee, Ducpyo Hong, Soocheol Kim, Jaeho Choi, Hyun-Mun Hong
 School of Electrical and Computer Engineering, Chungbuk National University

*Department of Electric and Electronic Engineering, Dong Hae University

ABSTRACT

Development of nonlinear loads and compensation instruments make PQ(Power Quality) problem into important issue. Few studies by signal processing and pattern classification as NN(Neural Network), Wavelet Transform, and Fuzzy present feature extraction. A lot of input features make not always good result and they are difficult to make realtime system. Thus, The dimensionality reduction is indispensable process. PCA(Principal Component Analysis) reduces high-dimensional input features onto a lower-dimensional subspace effectively. It will be useful to apply to realtime system and NN.

1. 서론

전력품질 문제는 종류가 다양하고 이에 따른 원인도 다양하다. 다양한 문제를 쉽게 자동으로 인식하고 그 원인까지 분석한다면 많은 전력품질 문제를 쉽게 해결이 가능할 것이다. 그러한 시스템을 위해 신경망이나 퍼지, 웨이블릿과 같은 많은 신호처리 기법에 대한 연구가 진행되어 왔으나 특히 신경망에 대한 연구는 적용가능성이 넓고 사람의 생각과 유사하다는 점에서 활발히 이루어지고 있다. 그러나 데이터 분류를 위해 중요한 문제는 광범위한 훈련 벡터의 입력이나 특징을 적절하게 선택하여 분류를 위한 특징으로 사용하는 것이다. 이러한 특징선택의 과정은 일반적으로 10개 이상의 특징을 갖는 실제 데이터에서는 기본적인 것이다. 이를 위해 변수를 기본 단위로 정규화하고 그 변이 값이 큰 변수를 선택하여 인식률을 높이는 과정이 중요하다^[1]. 그 중에서 패턴인식의 방법 중에 하나인 주성분 분석기법(PCA: Principle Component Analysis)은 입력을 선형적으로 조합하고 입력의 가치를 확인하여 입력 변수 선택을 가능하게 한다. 과도상태(Impulsive Transient, Oscillatory Transient), 잡음(Noise), 순시정전(Interruption)의 전력품질문제와

정상상태(Sine)에 대한 35가지의 특징들 중에서 주성분 분석기법을 통하여 중요도가 높은 특징을 선별하여 각각의 원인에 따른 주요특징을 분석하고 신경망의 특징으로 사용하기 위한 적절한 특징을 제시하였다.¹

2. 본론

2.1 주성분 분석기법

주성분 분석기법은 입력을 선형적으로 조합하고 입력의 가치를 확인하는 방법의 하나로 통신이론과 이미지처리에서는 Karhunen-Loeve 변환으로도 알려져 있다.

데이터 셀에 I 번째를 $x_i, i = 1, \dots, n$ 이라고 하자. 단 단위 벡터 u 로 사영한 후 u 는 데이터 셀의 최대화 된 변이가 될 것이다. 일반적으로 손실이 없다면 데이터 셀은 zero mean으로 (1)과 같이 가정될 수 있으며 u 로 x_i 의 사영은 (2)와 같이 내적으로 정의된다. u 는 단위 벡터로 (3)과 같이 나타낼 수 있으며 x_i 가 평균 영이므로 u 는 (4)와 같이 나타낼 수 있다.

$$\sum_{i=1}^n x_i = 0 \tag{1}$$

$$p_i = x_i \cdot u = x_i^T \cdot u = u^T \cdot x_i \tag{2}$$

$$\|u\| = \sqrt{u^T u} = 1 \tag{3}$$

$$\sum_{i=1}^n p_i = \sum_{i=1}^n u^T x_i = u^T \sum_{i=1}^n x_i = u^T \cdot 0 = 0 \tag{4}$$

p_i 의 제곱은 (5)와 같이 나타낼 수 있으므로 p_i 의 변이 사영은 (6)으로 나타낼 수 있다.

$$p_i^2 = (u^T x_i)(x_i^T u) = u^T (x_i x_i^T) u \tag{5}$$

$$\sigma_p^2(u) = \frac{1}{n} \sum_{i=1}^n p_i^2 \tag{6}$$

$$= \mathbf{u}^T \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{u} = \mathbf{u}^T \mathbf{R} \mathbf{u} \quad (6)$$

대칭행렬 \mathbf{R} 은 데이터 집합의 상관관계 행렬이다. 만약 데이터 집합이 zero mean이 아니면 \mathbf{R} 을 공분산행렬이라고 하고 (7)과 같으며, 이때 μ 는 $x_i, i=1, \dots, n$ 의 평균이다.

$$\frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \quad (7)$$

사영 변이 $\sigma_p^2(\mathbf{u})$ 는 라그랑지 곱셈기를 사용한 새로운 목적함수의 정의에 의해 정규화되는 단위 벡터상수에 따른다. (8)은 미분이 영이 되므로 (9)또는 (10)과 같다.

$$\mathbf{J} = \mathbf{u}^T \mathbf{R} \mathbf{u} + \lambda(1 - \mathbf{u}^T \mathbf{u}) \quad (8)$$

$$\nabla_{\mathbf{u}} \mathbf{J} = 2\mathbf{R} \mathbf{u} - 2\lambda \mathbf{u} = 0 \quad (9)$$

$$\mathbf{R} \mathbf{u} = \lambda \mathbf{u} \quad (10)$$

이 상태는 λ 가 상관관계 행렬 \mathbf{R} 의 고유값이고 \mathbf{u}_i 는 대응벡터임을 의미한다. 위의 stationary 상태에서 변이의 사영은 (11)과 같다.

$$\sigma_p^2(\mathbf{u}) = \mathbf{u}^T \mathbf{R} \mathbf{u} = \mathbf{u}^T \lambda \mathbf{u} = \lambda \mathbf{u}^T \mathbf{u} = \lambda \quad (11)$$

그러므로, 사영변이 $\lambda_p^2(\mathbf{u})$ 는 상관관계 벡터 \mathbf{R} 의 가장 큰 고유값과 같은 최대 값을 갖는다. 이것은 사영벡터 \mathbf{u} 와 대응 벡터와 같을 때 일어난다. 상관관계 벡터는 대칭이고 그것의 고유벡터들은 각각 수직이다. 주어진 벡터 \mathbf{x} 는 (12)와 같이 \mathbf{R} 의 n 개의 고유벡터를 사용해서 나타낼 수 있다.

$$\mathbf{x} = \sum_{i=0}^n q_i \mathbf{u}_i \quad (12)$$

만약 큰 변이를 가진 입력을 m 차원의 입력으로 차원 축소 한다면, \mathbf{x} 는 u_{m+1}, \dots, u_n 의 $n-m$ 항을 소거함으로써 $\hat{\mathbf{x}}$ 를 얻을 수 있다.

$$\hat{\mathbf{x}} = \sum_{i=1}^m p_i \mathbf{u}_i \quad (13)$$

그러므로 데이터 셀에서 차수축소를 하기위해서 상관관계 행렬과 고유벡터, 고유값을 먼저 찾아야한다.

주성분 분석기법은 그림 1에서와 같이 근사화 된 선이나 면에 수직인 오차 벡터의 길이를 최소화하는 데이터의 향으로 분석될 수 있다. P_i 는 i 번째 데이터 점 x_i 를 나타내고 M 은 모든 데이터의 점의 평균값이다. 직교법칙에 따라 M 에서 H_i 까지의 벡터는 P_i 에서 H_i 의 벡터와 수직이므로 (14)를 얻고, 모

든 데이터에 대해서 합을 취하면 (15)를 얻는다.

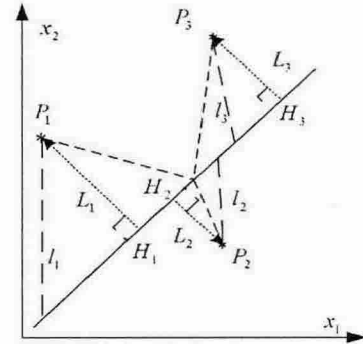


그림 1 주성분 분석기법의 기본개념
Fig. 1 Basic idea of PCA

$$\overline{MP_i^2} = \overline{MH_i^2} + \overline{P_iH_i^2} \quad (14)$$

$$\sum_i \overline{MP_i^2} = \sum_i \overline{MH_i^2} + \sum_i \overline{P_iH_i^2} \quad (15)$$

식 (15)에서 왼쪽 항은 일정하다. 오른쪽의 첫 번째 항은 주성분의 변이에 비례한다. 두 번째 항 $\sum_i L_i$ 은 근사화된 선에 직각방향의 거리의 합이다. 그러므로 주성분의 변이를 최대화하는 것은 점으로부터 선까지의 거리의 합인 $\sum_i L_i$ 을 최소화 하도록 하는 것이다. $\sum_i L_i$ 은 데이터 최적화 방법인 TLS(Total Least-Squares)를 통해 최소화 한다^[1-3].

2.2 특징추출

오랜 시간동안 지속적인 측정을 통해서 얻을 수 있는 전력품질의 파형들을 현실적으로 많이 얻기는 어려운 일이다. 그래서 몇 가지의 수집한 데이터와 논문 등의 자료를 통하여 실제와 같은 다양한 파형을 생성하였다. 특징을 추출하기 위해 과도상태 (Impulsive Transient, Oscillatory Transient), 잡음(Noise), 순시정전(Interruption)의 4가지의 전력품질문제 및 정상상태(Sine)파형을 각각 10가지씩을 생성하였다. 여기서 Oscillatory Transient중에서 back-to-back 커패시터 스위칭에 의해 발생하는 것을 Oscillatory1이라 하고, 커패시터 뱅크에 의해 발생하는 것을 Oscillatory2라 하였다^[4]. 이는 같은 Oscillatory지만 그 유형이 현저하게 다른데 따른 것이다. 생성한 데이터들을 바탕으로 통계적 처리방법인 평균(Mean), 표준편차(Standard deviation), 왜도(Skewness), 첨도(Kurtosis)와 파고율(Crest Factor), 실효값(RMS), 이산푸리에변환(DFT), 이산웨이블릿변환(DWT)을 통한 35가지의 특징량을 계산하였다. 계산결과를 주성분 분석기법을 통하여 사고유형에 따른 특징량 분석과 함께 전체적인 특징의 빈도를 분석하여 원하는 전력품질 문제에 적용 가능하도록 하였다.

2.3 실험 및 결과

5종류의 전력품질문제들에 대해 10가지 파형, 총 60가지의 파형에 대해 표 1과 같은 35가지의 특징을 계산하였다. 표 1에서의 D와 A는 웨이블릿변환의 세부신호(Detail)와 근사신호(Approximation)을 나타낸다.

표 1 번호에 따른 특징

Table 1 Feature names in each number

번호	특징	번호	특징	번호	특징
1	평균	13	D1의 표준편차	25	D3의 표준편차
2	표준편차	14	D1의 왜도	26	D3의 왜도
3	왜도	15	D1의 첨도	27	D3의 첨도
4	첨도	16	D1의 실효값	28	D3의 실효값
5	실효값	17	D1의 파고율	29	D3의 파고율
6	파고율	18	D2의 평균	30	D3의 평균
7	DFT 3차	19	D2의 표준편차	31	D3의 표준편차
8	DFT 5차	20	D2의 왜도	32	D3의 왜도
9	DFT 7차	21	D2의 첨도	33	D3의 첨도
10	DFT 9차	22	D2의 실효값	34	D3의 실효값
11	DFT 11차	23	D2의 파고율	35	D3의 파고율
12	DFT 13차	24	D3의 평균		

계산값을 주성분 분석기법을 통하여 고유벡터의 값을 구하였다. 고유벡터의 값이 클수록 중요도가 높은 특징량이므로 각파형에서 고유벡터의 가장 큰 값부터 여섯 번째까지의 중요벡터를 얻었다. 이 과정을 반복하여 각 5가지 전력품질문제인 과도상태 (Impulsive Transient, Oscillatory Transient), 잡음(Noise), 순시정전(Interruption), 정상상태(Sine)에 대한 결과는 표 2와 같다. 즉, Impulsive Transient: 19(D2의 표준편차)이고, 두 번째는 22(D2의 실효값), 세 번째는 26(D3의 왜도)순이다.

표 2 중요벡터의 빈도수

Table 2 Important vector frequency

전력품질 종류	중요벡터의 높은 빈도 순서
Impulsive Transient	19 > 22 > 25 > 28 > 13 = 16 = 28
Oscillatory Transient 1	19 > 22 > 13 = 16 = 25 = 28
Oscillatory Transient 2	28 > 19 > 25 > 13 > 16 > 22
Noise	27 > 6 > 21 > 33 > 1 = 3 = 23
Interruption	9 > 6 > 7 > 20 > 10 = 12 = 26
Sine	8 = 14 = 27 > 3 = 4 = 13 = 15

이러한 결과를 모두 더한 것이 그림 2와 같은 결과를 얻을 수 있다. 5가지의 전력품질 문제들의 특징의 중요도를 나타내는 결과로서 가장 중요벡터부터 나타내면, 19(D2의 표준편차), 13(D1의 표준편차), 26(D3의 왜도), 28(D3의 실효값)순이며 그 외에도 큰 값을 갖는 특징벡터로 6(파고율), 2(표준편차), 34(D3의 실효값)이다.

이를 10개의 파형으로 검증한 결과 90%의 인식률을 나타내었다.

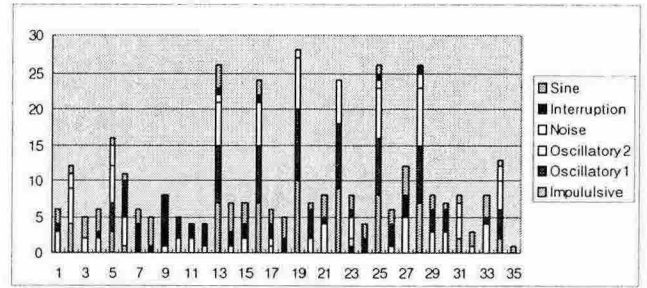


그림 2 모든 중요벡터의 합

Fig. 2 Summation of important vectors

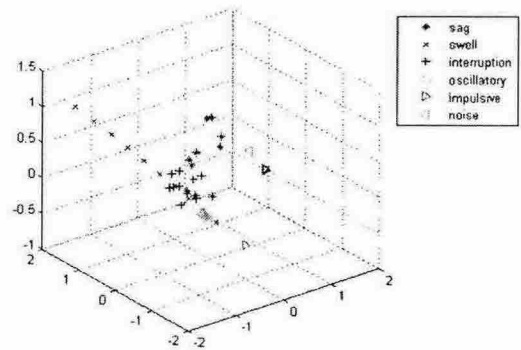


그림 3 주성분 분석기법의 인식결과

Fig. 3 Recognition result of PCA

3. 결론

전력품질문제를 분류하기 위해 중요한 문제는 다양한 훈련 벡터의 특징을 적절하게 선택하여 분류를 위한 특징으로 사용하는 것이므로 이를 위해 변수를 기본단위로 정규화하고 그 변이 값이 큰 변수를 선택하여 인식률을 높이는 과정이 매우 중요하다. 그 중에서 패턴인식의 방법 중에 하나인 주성분 분석기법은 입력을 선형적으로 조합하고 입력의 가치를 확인하여 입력 변수 선택을 가능하게 한다. 과도상태(Impulsive Transient, Oscillatory Transient), 잡음(Noise), 순시정전(Interruption)의 전력품질문제와 정상상태(Sine)에 대한 35가지의 특징들 중에서 주성분 분석기법을 통하여 중요도가 높은 특징을 선별한 결과 19(D2의 표준편차), 13(D1의 표준편차), 26(D3의 왜도), 28(D3의 실효값), 6(파고율), 2(표준편차), 34(D3의 실효값)의 순서를 중요도가 높다는 것을 알 수 있었다. 이러한 특징들은 차후의 신경망 연구에 밑거름이 될 수 있을 것이다.

참고 문헌

- [1] J.S.R. Jang, C.T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, 1997
- [2] 김상운, *패턴인식 및 학습*, 홍릉과학출판사, 2003
- [3] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, John Wiley & Sons, 2001
- [4] R.C. Dugan, M.F. Mcgranaghan, and H.W. Beaty, *Electrical Power Systems Quality*, McGraw-Hill, 1996