

## 무기체계의 신뢰성 향상을 위한 고장발생기간 중심의 대응분석 연구

송봉근<sup>1</sup>, 김근형<sup>2</sup>, 김용국<sup>2</sup>, 박승환<sup>1</sup>, 백준걸<sup>1\*</sup>

<sup>1</sup>고려대학교 산업경영공학과

<sup>2</sup>LIG넥스원 ILS연구센터

### The research of Correspondence Analysis centered on the Failure Period to improve the reliability of Weapon Systems

Bong-Geun Song<sup>1</sup>, Geun-Hyung Kim<sup>2</sup>, Young-Kuk Kim<sup>2</sup>, Seung Hwan Park<sup>1</sup>, Jun-Geol Baek<sup>1\*</sup>

<sup>1</sup>Department of Industrial Management Engineering, Korea University

<sup>2</sup>ILS(Integrated Logistics Support) R&D Lab, LIG Nex1

**요 약** 무기체계는 효율적인 전투준비태세를 갖추기 위해 개발단계의 신뢰성을 중요시하고 있다. 이미 제조업을 중심으로 다양한 분야에서 데이터 분석을 활용한 신뢰성 향상이 이루어지고 있다. 하지만 무기체계 개발단계는 보안의 중요성, 데이터의 부족 등으로 데이터 분석이 어려운 실정이다. 따라서 장기적인 무기체계 품질향상을 위해 전력화 이후의 장비 정보가 수집된 후속군수지원 데이터 분석을 수행하였다. 본 연구의 제안하는 방법론은 후속군수지원 데이터를 통해 목적변수인 고장발생기간을 중심으로 상관성 패턴을 파악하는 것이며, 절차는 다음과 같다. 첫 번째, 신뢰성에 영향을 미치는 주요 변수를 선택하고 고장발생기간을 중심으로 변수 간 상관성을 파악하였다. 두 번째, 범주형 데이터 특성을 갖는 데이터로부터 상관성 패턴을 파악하기 위해 대응분석 기법을 적용하여 분석을 수행하였다. 세 번째, 기여도와 표현력이 높은 범주들을 추출하고 시각화를 통해서 고장발생기간과 가장 관련이 높은 변수를 찾았다. 그리고 고장발생기간이 짧은 변수의 패턴을 선별하고 빈도분석을 통해서 신뢰성 저하 요인들을 파악하였다. 따라서 본 연구는 신무기 개발 시 신뢰성 저하 요인을 제거하여 군의 전투준비태세 강화에 도움이 될 것으로 기대한다.

**Abstract** Weapon systems require reliability in the development phase for efficient combat readiness. Improved reliability in various manufacturing processes have been achieved using data analysis. However, data analysis in the development phase is difficult due to problems such as the lack of data, high cost, and the importance of security. Therefore, Post Logistics Support (PLS) data collected following integration is analyzed for long-term quality improvement of weapon systems. In this study, we propose a methodology for examining the correlation between the failure rate and PLS data as follows: First, key variables affecting reliability were identified the correlation between variables on the failure rate examined. Second, corresponding analysis was conducted for determining the correlation between patterns of categorical data. Third, extract categories with the higher contribution and quality of representation, and find the highest variable correlated with failure period through visualization. Then, after selecting patterns which have shorter failure period, the cause of decreased reliability was confirmed through frequency analysis. This study will contribute to improving reliability when developing new weapon systems and will help to strengthen the combat readiness of military.

**Keywords** : Big Data Analysis, Correspondence Analysis, Exploratory Data Analysis, Failure Period, Post-Logistics Support, weapon systems

\*Corresponding Author : Jun-Geol Baek(Korea Univ.)

Tel: +82-2-3290-3396 email: jungeol@korea.ac.kr

Received August 26, 2016

Accepted October 7, 2016

Revised (1st September 26, 2016, 2nd October 6, 2016)

Published October 31, 2016

## 1. 서론

정보기술의 발달로 인해서 다양한 분야의 데이터가 급증하고 있다. 이에 따라 최근 제조업에서는 미래 제조산업의 변화를 주도하는 기술로써 빅 데이터 분석에 높은 관심을 갖고 있다. 특히 빅 데이터 분석을 활용한 공정 프로세스 개선은 제조업의 품질향상에 뛰어난 성과를 얻고 있다. 이러한 추세와 맞물려 국방 분야에서도 빅 데이터 분석을 도입하기 위해 다양한 연구를 수행하고 있다[1][2][3][4].

특히 군수분야는 군에서 빅 데이터 분석을 통해 비약적인 발전을 이끌어낼 수 있는 분야 중 하나로써 인식되고 있다. 현재 방대한 정보가 각종 군수 관련 정보체계를 통해서 획득되고 있다. 하지만 다양한 정보들이 유의하게 활용되지 못하고 사장되고 있는 실정이다[5].

현재 우리 군의 무기체계는 획득 및 운용/유지까지 단계별로 데이터가 수집되고 있다. 하지만 무기체계의 획득 및 운용/유지 단계별로 수행기관이 변경되기 때문에 업무 연계성이 미흡하다. 이와 같이 제대로 된 데이터 수집과 활용이 어려운 실정이며, 이는 무기체계 품질향상을 저해하는 요인이 된다[6].

무기체계는 품질향상을 통해 가동률을 높여 군의 전투준비태세 유지하며 유사시 그 성능을 발휘해야한다. 전투준비태세에 있어서 중요한 요소 중 하나는 우수한 품질의 무기를 만들고 의도하는 기간 동안 고장 없이 요구된 기능을 수행하는 것이다. 이를 위해서 무기체계 개발단계의 신뢰성 확보가 가장 먼저 고려되어야 한다. 신뢰성 저하는 잔 고장, 짧은 수명 및 정비 비용 증가의 원인이 되며 효율적으로 전투준비태세를 유지할 수 없게 된다[7].

하지만 무기체계는 높은 단위 비용, 장기간의 개발기간, 장소의 제한 등으로 많은 테스트 수행이 어렵다. 또한 무기체계 개발 단계에서는 개발 프로세스를 거치면서 단계적인 개선이 이루어지기 때문에 축적되는 데이터의 크기가 작고 동질성 있는 대량의 데이터 수집이 어렵다. 특히 무기개발은 보안의 중요성 때문에 개발단계에서 데이터 확보조차 어려운 실정이다. 따라서 현재 시험평가에서는 명중률과 같은 단순한 정량적 척도가 평가의 기준이 되고 있다. 즉, 무기체계 개발단계의 데이터 수준으로는 연구를 수행하는데 한계가 있다[8][9].

하지만 후속군수지원 단계의 데이터 활용은 이러한

무기체계 개발단계의 문제점을 개선할 수 있다. 후속군수지원은 전력화 이후 획득한 장비에 대한 야전 운용 및 정비의 모든 군수지원 과정을 의미한다. 후속군수지원 단계는 야전 운용을 통해 발생하는 결함에 대한 정비시간, 부품교체 이력, 결함원인 등의 정보 획득이 가능하다. 획득된 정보는 통합 데이터베이스에 구축되며 이를 이용하여 군수지원요소 최산화, 설계변경요소의 도출, 수리부속품 예측 등의 업무를 위한 분석을 수행할 수 있다[10].

후속군수지원 데이터베이스에는 무기체계를 효율적으로 운용하기 위한 활동들이 기록되어 있다. 따라서 후속군수지원 데이터는 전력화 후 무기체계들의 상태 정보를 제공하며 유의한 정보 획득이 가능하다. 이와 관련해서 정비 및 고장관련 데이터를 이용한 무기체계의 품질 및 신뢰도 향상 연구들이 수행되었다.

Kim 등(2015)은 장비정비 정보시스템의 정비이력 데이터, 업체가 관리하는 업체 보증수리이력 자료, 사용자 불만 정보를 통합하여 무기체계의 품질 개선을 위한 연구 방법을 제안하였다. Chung 등(2012)은 K계열 무기체계의 실 고장 데이터를 수집하고 이를 분석하는 과정을 통해서 무기체계의 신뢰성 평가 방법을 제안하는 연구를 수행하였다[11][12].

이처럼 최근 데이터를 이용한 다양한 연구들이 진행되고 있지만, 평균 및 빈도와 같은 단순 통계량 정보만을 사용한 경우가 대부분이었다. 단순 통계량은 데이터 전체를 설명하는데 효과적이지만 세부적인 정보를 살펴보기에는 적합하지 않기 때문에 고차원의 복잡한 데이터 정보를 이해하는데 어려움이 있다.

후속군수지원 데이터는 고차원 범주형 데이터의 특성을 갖는다. 따라서 이러한 특성의 데이터를 위한 분석기법이 필요하며, 본 연구에서는 데이터의 정보 추출을 위해서 탐색적 자료 분석(Exploratory Data Analysis) 기법 중 하나인 대응분석(Correspondence Analysis)이 활용되었다. 탐색적 자료 분석은 데이터에 내재하는 구조적 관계 및 특징 파악을 목적으로 한다. 또한 탐색적 자료 분석은 모델이 아닌 있는 그대로의 데이터를 통해서 새로운 통찰을 얻을 수 있는 장점이 있다[13].

본 연구를 위해서 후속군수지원 단계의 데이터로부터 개발단계의 신뢰성에 영향을 미치는 변수들을 추출하였다. 그리고 대응분석을 이용하여 변수 간 상관성 정보를 기반으로 유의한 정보를 추출하는 분석을 수행하였고,

무기체계 개발단계의 신뢰성 향상을 위한 정보를 제안하는 연구를 수행하였다.

본 논문의 구성은 다음과 같다. 1장은 연구의 배경 및 목적을 소개하였다. 2장은 대응분석을 활용한 데이터 분석 방법론을 제시하고 그에 대한 분석 및 결과를 설명하였다. 마지막 3장은 본 연구에 대한 결론을 작성하였다.

## 2. 본론

### 2.1 대응분석

본 연구는 고차원 범주형 데이터의 특성을 갖는 후속 군수지원 데이터를 분석하기 위해서 대응분석을 사용하였다. 대응분석은 탐색적 자료 분석 기법 중 하나로써 다음과 같은 장점이 있다.

대응분석은 두 범주형 변수의 상관관계 유·무뿐만 아니라, 분석 결과의 저차원 시각화를 통해서 데이터에 내재하는 패턴을 설명할 수 있다. 또한 대응분석은 탐색적 자료 분석을 위한 다른 기법들 보다 상관관계 특성 파악에 장점이 있다. 대표적인 기법으로 요인분석, 주성분분석, 군집분석이 있다. 요인분석은 공분산으로부터 변수들의 공통 요인들을 추출하며, 주성분분석은 데이터의 분산을 가장 잘 설명하는 주성분을 변수들의 선형조합으로 표현한다. 그리고 군집분석은 변수들 간의 유사도를 기반으로 분석을 수행한다. 세 가지 모두 변수 자체의 정보를 이용해서 분석을 수행한다. 하지만 대응분석은 변수 간 상관관계 정보뿐만 아니라 변수가 갖는 범주들이 어떤 패턴으로 관련되어 있는지 알 수 있는 장점이 있다. 따라서 대응분석은 대량의 범주형 데이터 특성을 갖는 후속군수지원 데이터로부터 내재하는 패턴을 파악하기 위해서 유용하게 활용될 수 있다[14].

대응분석의 계산 절차는 아래와 같이 수행되었다. 카이제곱 검정을 통해서 피어슨 잔차를 구한 뒤 특이값 분해를 수행하여 좌표 정보를 추출하고 시각화를 통해서 결과를 해석하였다.

#### 2.1.1 피어슨 잔차

카이제곱 검정은 교차표로 표현된 두 범주형 변수 간의 상관성을 파악하기 위해서 수행한다. 각 셀의 카이제곱값은 카이제곱 검정 과정을 통해서 계산되며 이를 피어슨 잔차(Pearson residuals)라고 부른다. 피어슨 잔차

는 카이제곱 검정의 가설에 대한 각 셀의 영향력을 나타내며 아래의 과정을 통해서 계산된다.

행 범주의 크기가  $p$ 이고, 열 범주의 크기가  $q$ 인 교차표  $F$ 의 전체 빈도수  $n$ 에 대해서 상대비율 행렬  $P$ 를 다음과 같이 정의한다.

$$P = \{f_{ij} / n\}, f_{ij} = \text{결합빈도}, i = 1, \dots, p; j = 1, \dots, q$$

$D_r$ 은  $(i,i)$ 요소가 행의 주변 비율( $r$ , row marginal proportion)인 대각행렬,  $D_c$ 는  $(j,j)$ 요소가 열의 주변 비율( $c$ , column marginal proportion)인 대각행렬로 정의하며 아래의 식(1), 식(2)로 계산된다[15].

$$Diag(D_r) = r, \quad r = \sum_{j=1}^q p_{ij} \quad (1)$$

$$Diag(D_c) = c, \quad c = \sum_{i=1}^p p_{ij} \quad (2)$$

피어슨 카이제곱 테스트(Pearson's chi-square test)는 교차표의 열과 행의 분포에 대한 균일성을 검증하는 방법으로 사용한다. 식(3), 식(4)는 피어슨 카이제곱 통계량(Pearson chi-squared statistic)과 피어슨 잔차를 나타낸다. 식(5)는 전체 관성(Total Inertia)으로써 행과 열의 종속성을 나타내는 상관계수로 사용된다[15].

$$\chi^2 = n \times \text{trace}[D_r^{-1/2}(P - rc^T)D_c^{-1}(P - rc^T)^T D_r^{-1/2}] \quad (3)$$

$$R = D_r^{-1/2}(P - rc^T)D_c^{-1/2} \quad (4)$$

$$\chi^2/n = \text{trace}(R^T R) \quad (5)$$

또한 각 행·열 변수가 갖고 있는 범주들 간의 카이제곱 거리는 단면(Profile) 정보를 이용해서 구할 수 있다. 예를 들어  $M$ 차원 공간 안의 행 단면(row profile)에서  $i^{th}$  행과  $i^{*th}$  행 사이의 카이제곱 거리는 수식(6)을 통해 계산할 수 있다[15].

$$d_I^2(i, i^*) = \sum_{m=1}^M (f_{im} - f_{i^*m})^2 = \sum_{j=1}^I \frac{1}{p_{+j}} \left( \frac{p_{ij}}{p_{i+}} - \frac{p_{i^*j}}{p_{i^*+}} \right) \quad (6)$$

#### 2.1.2 특이값 분해

대응분석은 행 변수와 열 변수 사이의 거리정보를 이용해서 저차원 공간의 그래프로 표현 가능한 장점을 갖고 있다. 이를 위해서 교차표의 범주들을 저차원 공간상의 점으로 표현할 수 있도록 좌표를 계산할 필요가 있다.

좌표는 식(4)를 통해 계산된  $R$ 에 대해 특이값 분해(Singular Value Decomposition)를 적용하여 쉽게 얻을 수 있으며 계산은 식(7)과 같다. 행과 열의 주좌표는 식(8), 식(9)로 계산되며 행과 열의 표준좌표는 식(10), 식(11)로 계산된다[15].

$$SVD(R) = U\Delta V^T \quad (7)$$

$$\text{principal coordinate of the row} = D_r^{-1/2}UD_\lambda \quad (8)$$

$$\text{principal coordinate of the column} = D_c^{-1/2}VD_\lambda \quad (9)$$

$$\text{standard coordinate of the row} = D_r^{-1/2}U \quad (10)$$

$$\text{standard coordinate of the column} = D_c^{-1/2}V \quad (11)$$

$D_\lambda$ 는 내림차순 정렬된 특이값(Singular Value,  $\Delta$ )의 대각행렬이며 고유값(Eigenvalues,  $\lambda = D_\lambda^2$ )을 구할 수 있다.

### 2.1.3 시각화

시각화는 여러 가지 의미 있는 분석 결과를 추출, 요약 및 설명하기 위해서 사용되어 진다. 따라서 시각화는 교차표의 행 범주와 열 범주를 좌표공간에 표현하여 대응분석 결과의 이해를 돕는다.

대응분석의 행과 열을 시각화하는 행렬도(Biplot)는 대칭그림(Symmetric plot)과 비대칭그림(Asymmetric plot)이 존재한다. 대칭그림은 주좌표(Principal coordinates)를 통해서 범주들을 시각적으로 보여준다. 그리고 비대칭그림은 표준좌표(Standard coordinates)로 표현된 행(또는 열) 벡터 공간에 주좌표로 표현되는 열(또는 행)의 단면 정보를 표현한 그림이다. 대칭그림은 행간 거리와 열간 거리의 정보는 대칭그림을 통해서 해석할 수 있다. 하지만 행과 열 사이의 거리에 대해서 해석을 제공하지 못하는 단점이 존재한다. 반면 비대칭 그림은 시각적으로 행-열의 관계 해석이 가능하기 때문에 본 연구는 비대칭그림을 사용해서 대응분석 결과를 해석하였다.

대응분석 해석 시 주의해야 할 점은 범주들의 거리가 상대적인 정보를 반영한다는 것이다. 주좌표로 표현된 두 범주가 벡터의 방향과 동일한 방향을 갖는다면, 원점으로부터 더 멀리 있는 범주에서 해당 벡터의 특성이 상대적으로 더 잘 나타난다. 또한 대응분석은 고차원의 정보를 저차원 공간에 표현하기 때문에 정보의 손실이 발생한다. 예를 들어  $10 \times 10$ 의 교차표에 대해서 완전하게 해당 정보를 표현하기 위해서는  $9(n-1)$ 차원이 필요하다.

따라서 대응분석을 해석할 때 표현력(Quality of representation)과 기여도(Contribution) 정보를 활용하면 유의한 해석이 가능하다. 아래의 식은 행과 열의 표현력과 기여도의 계산식을 나타낸다[16][17].

$$r * R.coord^2 / \lambda \quad (12)$$

$$c * C.coord^2 / \lambda \quad (13)$$

$$(\text{row coordinate})^2 / d^2(\text{row}, \text{average profile}) \quad (14)$$

$$(\text{col coordinate})^2 / d^2(\text{col}, \text{average profile}) \quad (15)$$

## 2.2 제안하는 방법론

본 연구는 무기체계 개발단계의 신뢰성에 영향을 미치는 변수를 추출하고 목적 변수와의 상관성을 파악하는데 중점을 두었다. 목적변수는 제조일부터 고장점수일까지의 기간을 의미하는 고장발생기간으로 설정하였다. 그리고 기간에 대한 상관성을 파악하기 위해서 대응분석을 적용하였다.

Fig. 1은 본 연구에서 제안하는 분석 절차를 설명한다. 분석은 후속군수지원 단계에서 수집된 데이터를 이용하여 수행되었고, 절차는 ‘데이터 수집(Database), 전처리(Pre-processing), 대응분석(Correspondence Analysis), 패턴분석(Pattern Analysis)’로 구성되어 있다.

### 2.2.1 전처리

후속군수지원 단계에는 무기체계의 운용이력, 고장이력, 정비이력, 개선정보에 대한 자료를 바탕으로 통합 데이터베이스가 구축되어 있다. 본 연구에서는 신뢰도에 영향을 미치는 원인 분석을 위해 필드결함 데이터로 분석을 수행했다.

필드결함 데이터는 분석을 위해서 데이터 탐색(Data Explore) 및 전처리(Data Pre-processing) 과정이 수행된다. 데이터 탐색 단계는 데이터의 특성을 파악하는 과정으로서 데이터의 전반적인 형태를 조사하고 분석에 필요한 변수 선택을 수행하였다. 전처리 과정은 데이터를 변환하는 작업으로써 잡음 데이터 제거 및 중복데이터 제거를 통해서 전처리 과정을 수행하였다. 잡음 데이터 제거는 결측값 및 불필요한 범주 수준의 제거를 통해 수행되었고, 이 과정을 통해서 데이터의 품질을 향상시킬 수 있다. 그리고 중복 데이터는 불필요한 저장 공간을 발생시키기 때문에 중복 데이터의 탐지 및 제거 과정을 수행하였다[18].

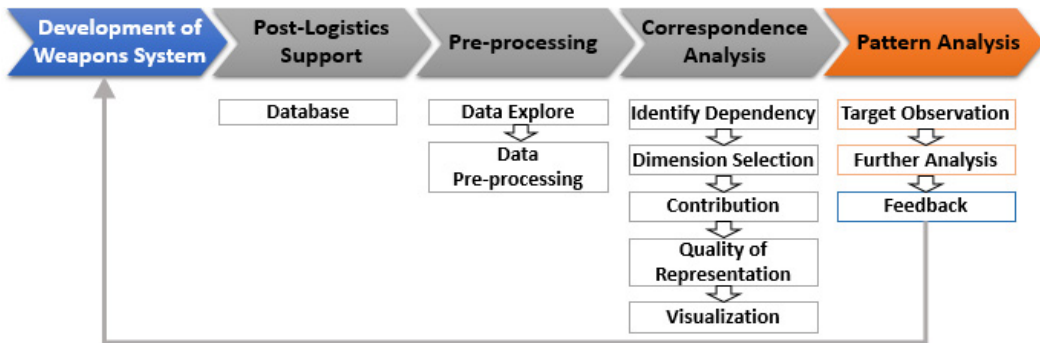


Fig. 1. Research Framework

### 2.2.2 대응분석

본 연구에서 수행한 대응분석 단계는 종속성 파악 (Identify Dependency), 분석 차원 선택(Dimension Selection), 기여도(Contribution) 및 표현력(Quality of Representation) 확인, 시각화(Visualization)의 단계로 구성된다.

대응분석을 위한 첫 번째 단계에서는 교차표의 행과 열 사이에 종속성(Dependency)이 존재하는지 살펴보았다. 종속성은 총 관성(Total Inertia)을 통해서 파악 가능하다. 총 관성은 교차표 테이블에 의해서 설명되는 전체 정보량을 의미하며 식(5)로 계산된다. 총 관성의 제곱근은 두 변수의 상관계수로 사용되며 0.2보다 큰 값을 나타내면 교차표의 행과 열 사이에 유의한 상관관계가 있는 것으로 해석할 수 있다[16].

두 번째 단계에서는 대응분석 결과를 표현하는 차원의 수를 결정하였다. 즉, 연관성의 패턴을 잘 표현하는 축의 개수를 선택하는 것과 같다. 많은 차원을 선택할수록 설명되는 데이터의 변동성은 증가하지만 해석을 복잡하게 만드는 트레이드오프(Trade-off)가 존재한다. 각 차원은 관성으로 설명 가능하다. 특이값의 제곱이 주성분 관성(‘principal inertia’ or ‘eigenvalue’)이 되며, 주성분 관성의 합은 총 관성으로 나타낸다. 총 관성에 대한 각 차원의 주성분 관성 비율에 따라서 각 차원이 해당 자료를 얼마나 설명하는지 나타낼 수 있다[19].

세 번째 단계에서는 기여도와 표현력을 고려하였다. 기여도는 각 축에 대해서 설명되는 범주들의 관성 비율을 나타낸다. 예를 들어 1차원 축에 대한 기여도 값이 높은 범주는 1차원 축에 대한 설명력이 높다고 볼 수 있다. 기여도는 범주의 수가 많은 경우 유용하게 사용될 수 있다. 기여도가 낮은 범주는 제거하고 설명력이 높은 범주

만을 이용해서 좌표공간의 속성을 해석할 수 있다. 기여도는 식(12), 식(13)으로 계산되며 각각 교차표의 행 기여도, 열 기여도에 대한 식을 나타낸다.

대응분석은 고차원 데이터를 저차원 공간으로 사상하므로 모든 정보들이 유의하게 표현되지 못한다. 각 범주들은 선택된 저차원 공간에서 잘 설명되는 범주가 있고 다른 차원 공간에서 잘 설명되는 범주가 있다. 따라서 저차원 공간에서는 우수한 표현력을 갖는 범주를 중심으로 결과를 해석할 필요가 있다. 표현력은 식(14), 식(15)을 통해서 계산되며 각각 행과 열의 표현력을 나타낸다. 본 연구는 기여도가 평균보다 높고 2차원 공간에서의 표현력이 우수한 범주들을 선택하여 결과 해석을 수행하였다[20].

대응분석의 마지막 단계는 시각적인 해석을 수행한다. 시각화는 여러 가지 의미 있는 분석 결과를 추출, 요약 및 설명하기 위해서 사용되어 진다. 본 연구에서는 두 변수에 내재하는 범주들의 해석을 위해 비대칭그림을 활용하였다. 비대칭그림은 두 변수에 대한 산점도를 나타낼 수 있으며 두 변수 간 상관성 패턴을 파악할 수 있는 장점이 있다.

### 2.2.3 패턴 분석

본 연구는 후속군수지원 단계에서 신뢰성에 영향을 미치는 주요 변수들을 추출하여 분석을 수행하였다. 목적변수는 고장발생기간으로 설정하였기 때문에 분석을 통해서 기간이 긴 특성 및 짧은 특성에 대한 패턴을 발견할 수 있다. 기간에 대한 정보를 추출하고 속성을 분석하는 것은 무기체계 개발단계의 유용한 피드백 정보로서 활용될 수 있다. 특히 기간이 짧은 정보는 해당 무기의 품질에 문제가 있음을 나타내며 전투준비태세의 주요 저해 요인으로 볼 수 있다. 따라서 본 연구는 대응

분석을 통해서 기간이 짧은 요소들과 상관관계가 높은 정보를 추출하고 빈도분석을 활용하여 정보의 속성을 파악하였다.

### 2.3 실험 결과

#### 2.3.1 전처리

수집된 데이터는 후속군수지원 단계에서 생성된 자료이다. 데이터는 무기체계의 수리 요청에 대한 정보를 포함하고 있으며 2011년부터 2014년까지 수집된 데이터로 구성되어 있다. 날짜, 부대, 무기체계, 결함원인, 조치사항, 금액 등의 정보들이 포함되어 있으며 대부분 범주형 변수로 구성되어 있다.

분석 변수 선정은 결측값 수와 범주의 수준 수(number of categorical levels)를 선택 기준으로 선정하였고 최종적으로 전문가의 의견을 종합하였다. 그 결과 Table. 1과 같이 5개의 변수가 선택되었다.

Table 1. Table title

Variables	Type	Levels	Missing value
Failure Period	Continuous	.	0
Military	Categorical	22	1040
weapon systems		96	116
Cause of Failure		6	63
TUM		4	0

지원부대(Military)와 무기체계(weapon systems)는 각각 수리요청 부대와 고장이 발생한 무기체계 종류를 나타낸다. TUM은 탑(Top), 유닛(Unit), 모듈(Module)을 의미하며 무기체계가 어떤 하위 부품 구조에서 고장이 발생했는지 나타낸다. 고장발생기간은 해당 무기체계의 제조일 부터 고장 접수일까지 기간을 나타낸다. 대응분석은 두 범주형 변수에 대한 분석을 수행한다. 따라서 목적변수인 고장발생기간은 범주형 변수로 변환 후 사용하였다. 고장발생기간은 이상값(Outlier)이 다수 존재하는 분포를 갖고 있기 때문에 동일 빈도 이산화(Equal Frequency Discretization)를 수행하여 범주형 변수로 변환하였다. 그 결과 고장발생기간은 세 개의 그룹으로 나누었다. 각 그룹은 기간이 짧은 순서대로 ‘FP1, FP2, FP3’로 명명하였다. ‘FP1’은 제조일 부터 1년 이내에 고장이 발생한 무기체계를 의미하며, ‘FP2’는 1년 이후 부터 2년 이내에 고장이 발생한 무기체계, ‘FP3’는 그 외에 제조일이 오래된 제품에 대한 무기체계를 포함한다[21].

이후 동일한 점수번호를 갖는 중복데이터는 제거하고 각 변수에서 범주의 수준수가 40개 이하인 데이터를 잡음 데이터로서 제거하였다. 전처리 과정을 거친 최종 분석 데이터는 지원부대(M1, M2, ..., M13), 무기체계(W1, W2, ..., W35), 불량원인(기타, 설계불량, 운용미숙, 자재불량, 작업불량), TUM(Top, Unit, Module)으로 선정했다.

#### 2.3.2 대응분석

대응분석의 첫 번째 단계는 변수 간 종속성 정보를 파악하기 위해 상관계수 값을 계산하였다. Table. 2는 고장발생기간과 각 변수들 간의 상관계수 값을 보여준다. 결과에 의해서 지원부대, 무기체계, 불량원인이 고장발생기간과 유의한 상관관계가 있는 것으로 보이며 TUM은 0.14로 고장발생기간과 유의한 상관관계가 존재하지 않는 것으로 확인되었다. 따라서 본 연구는 <고장발생기간과 지원부대>, <고장발생기간과 무기체계>, <고장발생기간과 고장원인> 3가지 경우에 대해서 대응분석을 수행하였다.

Table 2. Correlation Coefficient

	Military	Weapon System	Cause of Fault	TUM
Failure Period	0.27	0.73	0.20	0.14

Fig. 2은 전체 관성에 대한 각 차원의 비율을 나타내는 막대 그래프이다. 가로축은 1차원과 2차원을, 세로축은 각 차원이 갖는 관성을 나타낸다. 그리고 막대그래프 안에는 각 차원이 갖는 관성 비율의 수치가 나타나 있다. 세 가지 경우에 대한 관성 비율의 합은 모두 100%이다. 따라서 각각 1·2차원으로 전체 관성을 하고 있음을 알 수 있다. 따라서 본 연구는 2차원 공간에서 대응분석을 수행하였다.

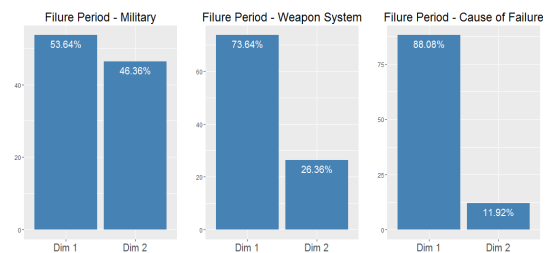


Fig. 2. Percentage of explained variances by dimension

대응분석 수행 시 고차원 공간의 범주들이 저차원에 유의하게 매핑되었는지 확인하기 위해서 기여도와 표현력을 고려하였다. Fig. 3, Fig. 4, Fig. 5은 <고장발생기간과 지원부대>, <고장발생기간과 무기체계>, <고장발생기간과 고장원인> 각각의 대응분석 수행 시 고장발생기간이 갖는 각 차원의 기여도를 나타낸 결과이며 점선은 평균 기여도를 나타낸다. 각 차원마다 높은 설명력을 갖는 범주는 다르지만 모든 범주들이 2차원 공간에서 평균 이상의 기여도를 보이는 것을 알 수 있다. 따라서 고장발생기간의 세 가지 범주는 저차원 공간의 속성 파악에 사용되었다.

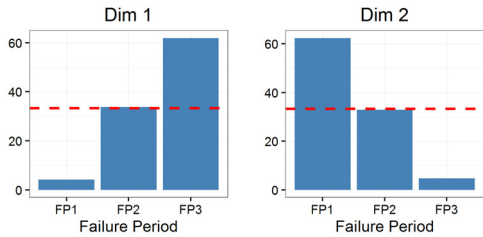


Fig. 3. Failure Period & Military

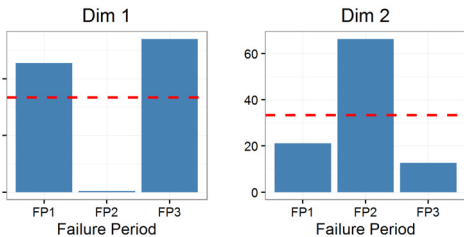


Fig. 4. Failure Period & weapon systems

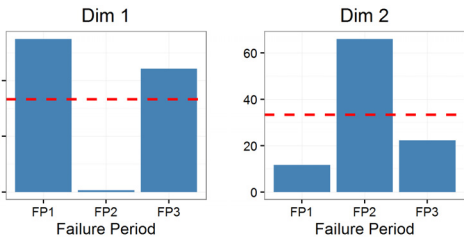


Fig. 5. Failure Period & Cause of Failure

Fig. 6은 세 가지 대응분석에 대해서 각 변수(지원부대, 무기체계, 불량원인)가 2차원에서 갖는 표현력을 나타낸다. 아래의 막대그래프를 통해서 각 변수의 모든 범주들이 100%의 표현력을 갖는 것을 확인할 수 있다. 이

는 Fig. 2에서 1-2차원 축이 관성을 100% 설명하고 있기 때문에 나온 결과로 해석할 수 있다. 따라서 지원부대, 무기체계, 불량원인이 갖고 있는 모든 범주를 활용하여 대응분석 결과를 해석하였다.

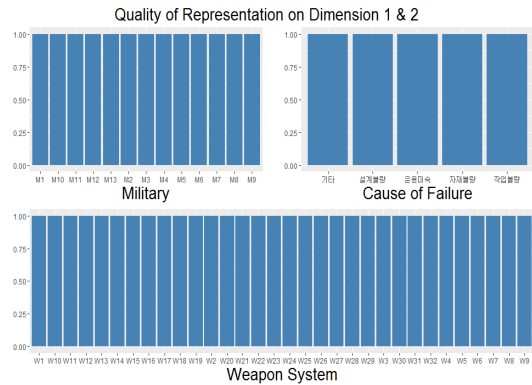


Fig. 6. Quality of Representation

Fig. 7, Fig. 8, Fig. 9는 <고장발생기간과 지원부대>, <고장발생기간과 무기체계>, <고장발생기간과 고장원인>에 대한 대응분석 시각화 결과이다. 각 그림의 화살표는 고장발생기간의 3가지 범주 FP1, FP2, FP3에 대한 벡터 방향을 의미한다. 벡터의 방향은 각 사분면에 대한 속성을 나타내며 주로 1차원 축을 기준으로 시간이 짧은 (FP1 or FP2) 속성과 시간이 긴(FP2 or FP3) 속성이 나누어진 것을 볼 수 있었다.

또한 Fig. 7, Fig. 8, Fig. 9에 나타난 벡터 주변의 범주들은 각각 지원부대, 무기체계, 고장원인의 범주들을 나타낸다. 분포된 범주들은 고장발생기간 범주의 화살표와 가깝고 원점에서 멀어질수록 화살표가 나타내는 정보의 상대빈도가 높은 특징을 보인다. 따라서 본 연구는 분포된 범주들과 각 벡터와의 유사도를 이용해서 그룹화를 수행하였다. 그룹화 결과는 Fig. 7, Fig. 8, Fig. 9에서 그룹 G1, G2, G3로 나타내었다.

각 그룹(G1,G2,G3)이 본 연구에서 알고자 하는 기간 정보 FP1, FP2, FP3의 빈도를 상대적으로 높게 나타내는 지 검증하기 위해 Table. 3을 구성하였다. 입력된 수치는 그룹 G1, G2, G3 각각에 대한 FP1, FP2, FP3의 상대빈도를 나타낸다. 고장원인은 FP2와 관련 있는 범주가 존재하지 않기 때문에 값이 계산되지 않았다. 하지만 나머지 그룹들은 각 그룹 G1, G2, G3에 대해서 해당 속성 FP1, FP2, FP3의 상대빈도가 높게 나타나는 것을 볼 수 있다.

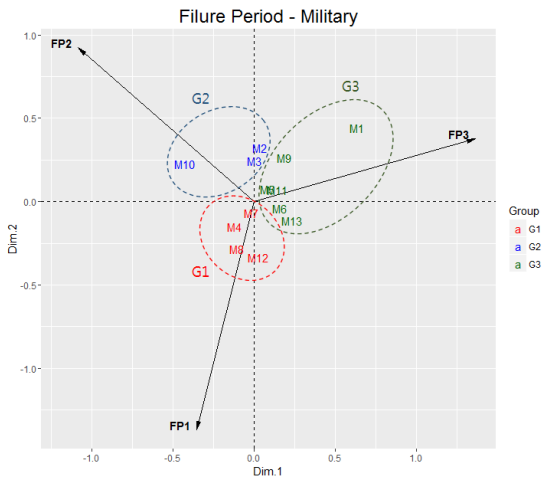


Fig. 7. Failure Period & Military

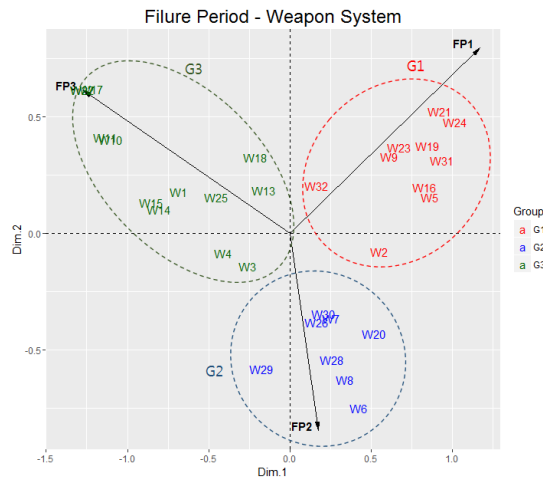


Fig. 8. Failure Period & weapon systems

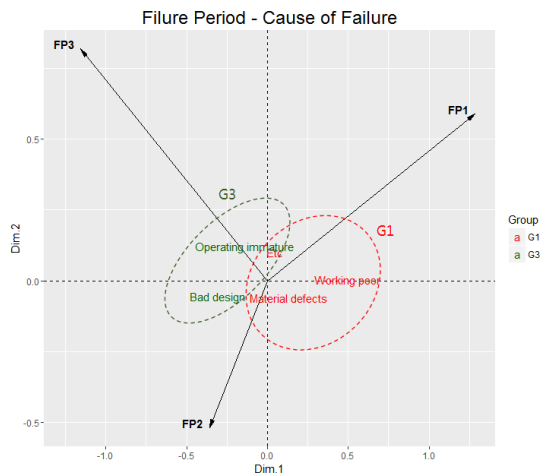


Fig. 9. Failure Period & Cause of Failure

세 가지 대응분석 중 <고장발생기간-무기체계>은 상대빈도 비율이 가장 높게 나타났고, <고장발생기간-고장원인>은 상대빈도 비율이 가장 낮게 나타났다. Table. 2의 결과와 비교했을 때 분석하고자 하는 두 변수의 상관계수 값이 높을 수록 대응분석 결과가 더 유의함을 볼 수 있다.

Table 3. Relative frequency of each group

Variables	Group	FP1	FP2	FP3
Military	G1	<b>0.43</b>	0.3	0.27
	G2	0.25	<b>0.49</b>	0.26
	G3	0.3	0.29	<b>0.42</b>
Weapon System	G1	<b>0.7</b>	0.25	0.05
	G2	0.2	<b>0.58</b>	0.22
	G3	0.15	0.2	<b>0.64</b>
Cause of Failure	G1	<b>0.39</b>	0.33	0.28
	G2	.	.	.
	G3	0.23	0.34	<b>0.43</b>

### 2.3.3 패턴 분석

본 연구에서는 고장발생기간과 관련된 특성을 파악하기 위해서 지원부대, 무기체계, 고장원인과 같은 변수들을 활용해서 대응분석을 수행하였다. 그 결과 무기체계와 고장발생기간에서 가장 상관성 있는 패턴이 존재함을 발견하였다. 이후 고장발생기간이 짧은 무기체계를 선별하고 속성을 파악하기 위한 분석을 수행하였다.

분석을 위해서 고장발생기간이 2년 이내인 무기체계를 선별하였다. 그 결과 Fig. 8의 FP1 벡터와 FP2 벡터 사이에 분포된 무기체계 범주들이 선택되었다. 그리고 각 무기체계의 특성은 고장원인, 지원부대, TUM의 빈도 분석을 통해서 파악하였다.

Table. 4은 고장발생기간이 2년 이내의 특성을 보이는 무기체계들에 대한 추가분석 결과를 나타낸다. 데이터의 특성상 각 변수마다 특정 범주에 빈도가 높게 형성되어있다. 결합원인은 자체불량에서, TUM은 Module에서 높은 빈도를 나타낸 반면 지원부대는 대체적으로 고른 분포를 보였다. 따라서 각 무기체계의 특성은 각 속성 별로 두 번째로 높은 상대빈도까지 고려해서 파악하였다.



**Table 4.** Attributes of weapon systems

weapon systems	Cause of Failure(%)		Military(%)		TUM(%)	
	1st	2nd	1st	2nd	1st	2nd
W23	Material defects (60)	Etc (18)	M5 (23)	M3 (15)	Module (97)	Unit (3)
W28	Material defects (64)	Etc (19)	M10 (99)	M11 (1)	Module (99)	Unit (1)
W24	Material defects (42)	Bad design (18)	M7 (40)	M6 (31)	Module (94)	Unit (6)
W2	Material defects (53)	Bad design (21)	M11 (33)	M4 (25)	Module (97)	Unit (3)
W21	Material defects (61)	Working poor (18)	M4 (53)	M5 (31)	Module (96)	Unit (4)
W5	Material defects (44)	Operating immature (21)	M10 (99)	M11 (1)	Top (63)	Module (25)
W31	Etc (30)	Material defects (27)	M8 (67)	M4 (29)	Module (81)	Unit (8)
W8	Bad design (47)	Etc (39)	M11 (100)	M1 (0)	Top (100)	Module (0)

W23, W28, W24, W2, W21은 지원부대의 특성이 상이하지만 공통적으로 자재불량의 불량원인과 Module에서 고장이 발생한 특성을 보인다. W5는 위의 무기체계와 다르게 TOP에서 발생한 고장 비율이 높으며, 불량원인의 두 번째 요인은 운용미숙의 특성을 보인다. W31, W8 무기체계는 불량원인에서 상이한 특성을 보인다. W31은 기타의 불량원인이 높으며 자재불량이 두 번째로 높은 불량원인 특성일 보인다. W8은 설계불량과 기타의 불량원인 특성이 높고 모두 TOP에서 불량이 발생한 것을 볼 수 있었다. 이를 활용하여 고장발생기간이 짧은 무기체계의 속성을 파악하고 유용한 정보로 활용될 수 있다.

### 3. 결론

본 연구는 무기체계 개발단계의 품질향상 기반을 마련하기 위한 연구를 제안하였다. 데이터는 후속군수지원 단계에서 생성되는 무기체계 필드 결합 데이터를 활용하였다. 획득된 데이터는 수기로 입력된 범주형 데이터의 특성을 갖고 있기 때문에 대응분석을 수행하였다. 대응분석의 시각적 해석은 단순 빈도를 통해서 알기 어려운

다양한 패턴을 쉽게 해석할 수 있는 장점을 갖는다. 따라서 본 연구는 고장발생기간을 중심으로 지원부대, 무기체계, 불량원인 변수들과 대응분석을 수행하였고, 시각화를 통해서 기간에 대한 정보의 패턴을 파악하였다.

고장이 빨리 발생하는 무기체계는 전투준비태세의 저하 요인이 되기 때문에 고장발생기간이 짧은 패턴을 중심으로 탐색하였다. 분석 결과는 무기체계에서 고장발생기간이 짧은 패턴과 가장 유의한 상관성이 존재함을 보였다. 그래서 2년 이내에 고장이 발생한 무기체계들에 대해 추가 분석을 수행하였다. 선별된 무기체계는 W2, W5, W8, W21, W23, W24, W28, W31이다. 그리고 속성은 불량원인, 지원부대, TUM 변수를 통해서 파악하였다. 그 결과 각 무기체계별로 속성을 통해서 집중적인 관리가 필요한 요소를 확인할 수 있었다. 따라서 이러한 데이터 분석은 고장 위험요소들을 식별하고 무기체계의 품질불량을 방지할 수 있을 것으로 기대한다.

추후 연구로는 보다 유의한 정보를 얻기 위해서 다양한 데이터를 수집하여 분석을 수행할 필요가 있다. 신뢰성에 영향을 주는 변수가 제한되어 있었기 때문에 주어진 변수 정보만을 이용해서 분석을 수행하였다. 하지만 데이터의 특성 및 연구자의 관심도에 따라서 다양하고 새로운 분석을 진행하게 된다면 더욱더 유용한 정보를 확보할 수 있을 것으로 기대한다.

### References

- [1] "Bigdata Monthly, BigData Trends and Issues", vol. 10, 2015
- [2] S.W. Kim, G.G. Kim, Bong-Kyu Yoon, "A Study on a Way to Utilize Big Data Analytics in the Defense Area", Korean Operations Research And Management Society, vol. 39, no. 2, pp. 1-19, 2014.  
DOI: <http://dx.doi.org/10.7737/JKORMS.2014.39.4.001>
- [3] K.S. Lee, "Construction and Application of the Defense Acquisition Database as a Big Data", Institute of Control, Robotics and Systems, pp. 125-137, 2012.
- [4] S.Y. Jeon, D.H. Lee, Manjae Bae, "Study on the Application Method of Munition's Quality Information based on Big Data", Korea Academia-Industrial cooperation Society, vol. 17, no. 6, pp. 315-325, 2016.
- [5] C.G. Ku, G.Y. Yoo, Y.H. Kim, J.H. Ahn, "Big data and Application of Logistics field", Korea Defense Industry Association, vol. 446, pp. 62-73, 2016.
- [6] J.Y. Moon, "The Total Life Cycle System Management(TLCSM) for weapon systems Suggestions", Korea Defense Industry Association, vol. 435, pp. 76-83,

2015.

- [7] H.E. Kim, B.S. Kang “Improve the reliability of machinery parts”, The Korean Society of Mechanical Engineers, vol. 51, no. 8, pp. 34-38, 2011.
- [8] D.H. Kim, “A Study on the Test & Evaluation Method of Guided Missiles based on Reliability Growth Management”, Kwangwoon University, 107, 2015.
- [9] G.H. Kim, Y.K. Kim, S.H. Park, “A framework for Quality Improvement in weapon systems using Post-Logistics Support Data”, Korea Academia-Industrial cooperation Society, vol. 17, no. 5, pp. 680-687, 2016.
- [10] T.G. Kim, G.B. Wi, N.C. Lee, “The Research on the Applying Method for the Life Cycle Sustainment Plan of the weapon systems”, Korea Institution of National Defense Development, 222, 2012.
- [11] H.G. Kim, S.M. Kwon, K.H. Cho, S.I. Sung, “Development of Quality Improvement Process based on the Maintenance Data of weapon systemss”, Defense Agency for Technology and Quality, vol. 43, no. 4, pp. 499, 2015.  
DOI: <http://dx.doi.org/10.7469/jksqm.2015.43.4.499>
- [12] I.H. Chung, H.Y. Lee, Y.I. Park, “Reliability Evaluation of weapon systems using Field Data: Focusing on Case Study of K-series weapon systems”, The Korean Society for Quality Management, vol. 40, no. 3, pp. 278, 2012.  
DOI: <http://dx.doi.org/10.7469/JKSQM.2012.40.3.278>
- [13] M.H. Huh, “Exploratory data analysis using R”, pp37-38, Freecademi, 2007.
- [14] Doey, Laura, Jessica Kurta, "Correspondence analysis applied to psychological research.", Tutorials in Quantitative Methods for Psychology vol. 7, no. 1, pp. 5-14, 2011.  
DOI: <http://dx.doi.org/10.20982/tqmp.07.1.p005>
- [15] Beh, Eric J. "Simple correspondence analysis using adjusted residuals.", Journal of Statistical Planning and Inference, vol. 142, no. 4, pp. 965-973, 2012.  
DOI: <http://dx.doi.org/10.1016/j.jspi.2011.11.004>
- [16] Bendixen, Mike. "A practical guide to the use of correspondence analysis in marketing research." Marketing Research On-Line 1.1, 16-36, 1996.
- [17] wHerman, Steve, "MCA+.", New Jersey: Bretton-Clark, 1991.
- [18] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3, 1996.
- [19] Alberti, Gianmarco, "An R script to facilitate Correspondence Analysis: a guide to the use and the interpretation of results from an archaeological perspective.", Archeologia e calcolatori 24, pp. 25-53, 2013.
- [20] Greenacre, Michael, “Correspondence analysis in practice”, CRC press, 87, 2007.  
DOI: <http://dx.doi.org/10.1201/9781420011234>
- [21] Muhlenbach, Fabrice, and Ricco Rakotomalala, "Discretization of continuous attributes.", Encyclopedia of Data Warehousing and Mining 1 , pp. 397-402, 2005.  
DOI: <http://dx.doi.org/10.4018/978-1-59140-557-3.ch076>

**송 봉 근(Bong-Geun Song)**

[준회원]



- 2014년 8월 : 연세대학교 정보통계학과 (이학학사)
- 2014년 9월 ~ 현재 : 고려대학교 산업경영공학과 석사과정

<관심분야>

Data mining applications

**김 근 형(Geun-Hyung Kim)**

[정회원]



- 2009년 8월 : 인하대학교 산업공학과 (공학학사)
- 2012년 2월 : 고려대학교 정보경영전문대학원 정보경영공학과 (공학석사)
- 2013년 1월 ~ 현재 : LIG넥스원 선임 연구원

<관심분야>

종합군수지원(ILS), Data mining, Scheduling

**김 용 국(Young-Kuk Kim)**

[정회원]



- 2005년 8월 : 상지대학교 산업공학과 (공학학사)
- 2007년 8월 : 한양대학교 일반대학원 산업공학과 (공학석사)
- 2007년 8월 ~ 2011년 12월 : 국방과학연구소
- 2011년 12월 ~ 현재 : LIG넥스원 선임 연구원

<관심분야>

종합군수지원(ILS), Data mining, SCM

**박 승 환**(Seung Hwan Park)

[정회원]



- 2008년 2월 : 고려대학교 전산학과 (이학학사)
- 2011년 2월 : 고려대학교 산업경영공학과 (공학석사)
- 2011년 3월 ~ 현재 : 고려대학교 산업경영공학과 박사과정

<관심분야>

Data analytics for manufacturing process  
Statistical process control for profile data  
Quality and yield management  
Data mining for business analytics

**백 준 겐**(Jun-Geol Baek)

[정회원]



- 1993년 2월 : 고려대학교 산업공학과 학사 (공학학사)
- 1995년 2월 : 고려대학교 산업공학과 석사 (공학석사)
- 2001년 2월 : 고려대학교 산업공학과 박사 (공학박사)
- 2001년 3월 ~ 2002년 2월 : 고려대학교 정보통신기술연구소 연구조교수
- 2002년 3월 ~ 2007년 2월 : 인덕대학 산업시스템경영학과 조교수
- 2007년 3월 ~ 2008년 8월 : 평운대학교 경영학부 조교수
- 2008년 9월 ~ 현재 : 고려대학교 산업경영공학과 교수

<관심분야>

Advanced process control, Intelligent fault detection,  
Data mining applications