

Machine Learning을 이용한 무기 체계(or 구성품) 고장 유형 식별

박연경*, 이해원, 김상문
LIG넥스원ILS연구센터

Identify the Failure Mode of Weapon System (or equipment) using Machine Learning

Yun-Kyung Park*, Hye-Won Lee, Sang-Moon Kim
ILS R&D Lab, LIG Nex1 Co., Ltd.

요약 무기 체계(or 구성품) 개발은 한정된 개발기간과 비용 등의 제한으로 시험 횟수가 많지 않아, 고장관련 축적된 데이터의 규모도 적다. 그러나 운용 중 발생한 고장 및 정비내역은 많은 부분 전산 데이터로 관리하고 있기 때문에 이를 활용한 무기 체계(or 구성품)의 고장원인 분석은 가능하다. 다만 다양한 무기체계의 고장 및 정비내역 작성 규격이 각 군 별, 업체별 상이하고, 고장 원인의 구체적 내역은 비정형 텍스트 데이터로 기술되어 있기 때문에 이를 분석하는데 어려움이 있었다. 그러나 오늘날 빅데이터 처리 기술과 기계학습(Machine Learning) 알고리즘의 발전, HW연산 능력의 개선과 맞물려, 상기와 같은 비정형 데이터를 처리 할 수 있는 여러 가지 방법들이 시도 되고 있으며, 주요한 연구 분야로 활발히 연구되고 있다. 본 논문에서는 국방 무기 체계(or 구성품)의 고장/정비 관련 비정형 데이터를 기계학습 기법 중 하나인 doc2vec을 적용하여 고장 사례 분석 방안에 대하여 제시한다.

Abstract The development of weapon systems (or components) is hindered by the number of tests due to the limited development period and cost, which reduces the scale of accumulated data related to failures. Nevertheless, because a large amount of failure data and maintenance details during the operational period are managed by computerized data, the cause of failure of weapon systems (or components) can be analyzed using the data. On the other hand, analyzing the failure and maintenance details of various weapon systems is difficult because of the variation among groups and companies, and details of the cause of failure are described as unstructured text data. Fortunately, the recent developments of big data processing technology, machine learning algorithm, and improved HW computation ability have supported major research into various methods for processing the above unstructured data. In this paper, unstructured data related to the failure / maintenance of defense weapon systems (or components) is presented by applying doc2vec, a machine learning technique, to analyze the failure cases.

Keywords : Machine learning, doc2vec, Clustering, Visualization, Failure mode, Weapon System

1. 서론

오늘날 데이터의 중요성이 강조 되고, 컴퓨팅 기술이 발전하면서 여러 분야에서 각종 데이터가 축적되고 있다 [1]. 이는 국방 무기체계의 정비 분야에서도 동일하게 나

타나는 현상이다. 군(軍)의 경우 주요 장비의 고장 및 운영 자료를 확보 할 수 있는 체계가 운영되고 있으며, 군수 업체의 경우 정비내역 관리 체계 등을 통하여 무기체계의 고장 징후, 원인, 수리내역, 수리 부품 등이 전산화 되어 축적되고 있다[2]. 그러나 국방 정비 관련 데이터의

*Corresponding Author : Yun-Kyung Park(LIG Nex1 Co., Ltd)

Tel: +82-31-288-9405 email: ykpark2002@lignex1.com

Received May 23, 2018

Accepted August 3, 2018

Revised (1st June 7, 2018, 2nd June 22, 2018)

Published August 31, 2018

양적 팽창은 진행되고 있으나, 방대한 데이터를 활용하여 의미 있는 정보를 추출하여 활용하는 방안은 미흡한 현실이다[3]. 기존 데이터 분석은 정형화된 데이터를 수집·정제하여 분석하는 방식[4]으로 비정형 정비 데이터의 활용에 한계가 있었다. 방대한 양의 비정형 데이터를 기존 데이터 분석 방식으로 활용하기 위해서는 분석 목적에 적합한 데이터 전처리(Pre-Processing) 과정이 필요로 하며, 이는 인력 및 시간, 비용의 소요가 필요함을 의미한다. 그러나 최근 이러한 비정형 데이터 활용의 한계를 극복할 수 있는 빅데이터(Big Data) 검색 및 기계학습(Machine Learning) 기법들이 개발되고 있으며[4], 일부 분야에서 이미 서비스되고, 상업화되고 있다.

무기 체계 후속지원분야(후속 정비지원 포함) 비용은 무기 체계 수명주기 비용의 60~70%를 차지하므로[5], 수명주기 비용 절감을 위해서는 무기 체계 개발 시 주요 고장에 대한 강건 설계를 고려하여 개발하는 것이 중요하다. 또한 고장을 예방/보수하기 위하여 주요 고장 품목에 대한 군수지원(Logistic Support) 요소의 고려 역시 중요하다.

국내에서 무기 체계 개발 시, 고장 강건 설계 및 군수지원요소 개발을 위하여 수행되고 있는 주요한 활동은 MIL-STD-1629A 기준의 FMECA (Failure Mode Effect and Critical Analysis)이다[6]. FMECA의 핵심인 고장 모드, 원인, 영향(Failure Mode, Cause, Effect)은 주로 설계자의 주관적이고 경험적 의견으로 선정되고 있으며, 이를 기본으로 고장에 대한 강건 설계, 고장 식별 방안, 정비 방안이 결정되어 개발되고 있다. 그러나 설계자의 주관적이고 경험적 의견을 지원할 수 있는 유사장비에 대한 고장 모드, 원인, 영향에 대한 체계적인 자료(Data)는 전무한 실정이다.

유사 장비에 대한 고장 모드, 원인, 영향 (Failure Mode Effect and Critical Analysis)에 대한 자료는 FMECA의 수행을 위해서도 중요할 뿐만 아니라, 개발 장비의 전력화 배치 및 실제 운용 시 발생할 수 있는 주요 고장에 대한 통찰력을 주며, 개발 장비의 핵심 구성품 파악 및 고장 강건 설계 대상 선정 시에도 활용될 수 있다. 다만, 지금까지 유사 장비에 대한 고장 모드, 원인, 영향(Failure Mode, Cause, Effect)에 대한 체계적인 종합이 어려운 이유는 자료의 비정형성(Text Document Data)으로 인한 분석의 제한을 원인으로 꼽을 수 있다[3].

최근 H/W (CPU, GPU)의 발전으로 연산속도의 증가와 인공 신경망(Neural Network) 기반의 비정형 텍스트 데이터 분석 방법(Glove, Word2Vec, doc2vec 등)은 컴퓨터를 활용하여 과거 산만하게 널려 있던 무기 체계 정비 문서(Text Document Data)의 종합적이고 체계적인 학습을 가능하게 하고 있다[1].

본 논문에서는 무기체계(or 구성품) 정비 문서(Text Document Data)에 인공 신경망 기반 비정형 데이터 분석 방법인 doc2vec을 적용하여 무기체계 별 대표 고장 사례를 분석하는 방법을 제시한다. 또한, 분석 결과 학습된 무기체계 별 벡터를 기준으로 유사 무기 체계를 식별하는 무기 체계의 군집화(Clustering)/시각화(Visualization) 결과를 제시한다.

본 논문의 구성은 다음과 같다. 관련 연구(2.1), 연구 프로세스, 데이터 수집 및 기계학습을 적용하는 방법 제시(2.1~2.6), 학습결과 평가(2.7), 기계학습 적용을 통한 무기체계의 대표 고장사례 도출 및 군집화/시각화 결과와 활용 방안 제시(2.8), 마지막으로 결론에서는 연구결과 요약 및 향후 연구 방향에 대하여 정리한다.

2. 본론

2.1 관련연구

2.1.1 데이터 벡터화(Word Embedding)

기계학습을 위한 비정형 텍스트 빅데이터의 처리 기술로 텍스트 데이터의 벡터화가 많이 적용되고 있다. 데이터 벡터화란 텍스트를 구성하는 단어를 수치화하는 것으로서, BOW(Bag of Words) 기법이 주로 사용된다. BOW는 문서에 있는 단어에 정수를 할당함으로써 여러 문서의 카테고리를 나누는 데 사용되며, 대표적인 모델로 Word2Vec이 있다. BOW 기법을 기반으로 데이터에 레이블을 부여하여 기계학습을 수행하는 지도학습(Supervised Learning), 데이터에 레이블을 부여하지 않고 데이터 특성을 통해 기계학습을 수행하는 비지도학습(Unsupervised Learning)이 가능하다[8-9].

2.1.2 Paragraph Vector

문단 벡터(Paragraph vectors)는 Word2Vec의 확장된 모습으로서 문단을 벡터화하는 것을 말하며, doc2vec으로도 불린다. 문단벡터는 Le and Mikolov (2014)에 의

하여 처음 제시되었으며, doc2vec에는 dbow 방식과 dmpv 방식이 있다. dbow는 단어의 순서를 고려하지 않고 데이터를 수치화 하는 방식이며, dmpv는 단어의 순서 및 문맥을 고려하여 데이터를 수치화하는 방식이다 (Le and Mikolov , 2014)[10].

2.1.3 t-SNE

t-SNE(t-Distribution Stochastic Neighbor Embedding)의 아이디어는 데이터 포인트 사이의 거리를 가장 잘 보존하는 2차원표현을 찾는 것이다. 각 데이터 포인트를 2차원에 무작위로 표현한 후 원본 특성 공간에서 가까운 포인트는 가깝게, 멀리 떨어진 포인트는 멀어지게 조정한다. t-SNE는 멀리 떨어진 포인트와 거리를 보존하는 것보다는 가까이 있는 포인트에 더 많은 가중치를 두어 표현하는 특성을 가진다[7].

2.1.4 DBSCAN

군집화(Clustering) 알고리즘인 DBSCAN(density based spatial clustering of applications with noise)은 주요 군집화 알고리즘과 다르게 군집(Cluster)의 개수를 미리 지정 할 필요가 없다는 장점을 지닌다. DBSCAN의 두 개의 매개 변수는 최소 군집 수량 (min samples) 와 eps 거리를 설정하면, 해당 조건에 만족하는 군집의 개수를 식별하고, 어떤 군집에도 속하지 않은 포인트 (Noise)도 식별해 제시 한다[7].

2.2 연구 프로세스

본 연구의 전체 적인 과정은 Fig. 1.과 같다. 1단계에서는 모 방산업체에서 약 10년간 기록해온 무기체계 정비 문서를 수집하였다. 2단계에서는 수집한 정비문서를 분석하기 위하여 정비 문서에서 특수문자, 단일단어, 조사, 다빈도 무의미 단어 등을 처리하는 전처리(Pre-Processing)를 수행 하였다. 3단계에서는 수정된 정비문서 데이터에 기계학습 기법인 doc2vec을 적용하여 벡터를 학습하였다. 4단계에서는 기계학습 결과의 타당성을 확인하기 위하여 적중률을 정의 하였으며, 이를 기준으로 학습 성능을 평가하였다. 마지막으로 5단계에서는 학습된 무기체계 ID 벡터와 정비 문서 간 Cosine Similarity를 통해 대표 문서를 식별하였고, 고장 정비 내역 기준 무기체계 간의 유사성을 직관적으로 확인할 수 있도록 군집화 및 시각화를 수행하였다. 또한 이를 활용할 수 있는 방안을 제

안 하였다.

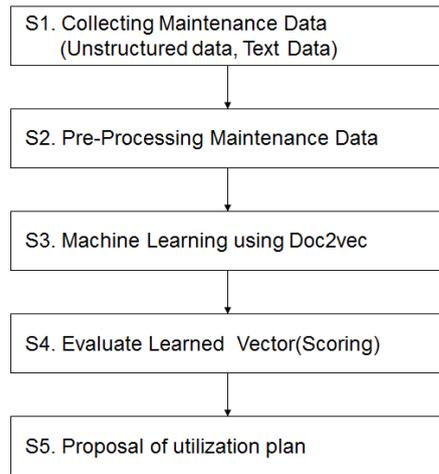


Fig. 1. Machine Learning Process on Weapon System (or Equipment) Maintenance Data

2.3 데이터수집

본 연구는 doc2vec 적용을 위해 2007년 1월부터 2017년 12월까지 모 방산업체에서 작성된 정비문서를 수집하였으며, 해당 무기체계의 정비 빈도가 50건 이상인 32개 무기체계(or 구성품)에 대한 정비문서 6,382 건을 선별하였다. 수집된 데이터는 양산 후 운용 3년 이내 (A/S 기간 내) 초기 고장 발생에 대한 원인 및 정비 내역으로서, 무기체계(or 구성품)의 초기 고장 원인 분석에 유용하게 활용 될 수 있다.

수집된 정비 문서에는 Table 1.과 같이 장비ID, 고장 징후, 고장 원인, 수리 내역, 수리 부품이 포함되어 있고, 자료의 형태는 비정형 텍스트 데이터이다. Table 2.는 정비 문서 데이터의 일부 예시이다.

Table 1. Data Collection

ID.	Details
Data Type	Text Documents
Num of Docs	6,382 ('07.1~17.12)
Num of Equip	32
Contents	<ul style="list-style-type: none"> • Equip_ID • Symptom • Cause • Repair • Parts

Table 7. doc2vec Tag Configuration

Tagged_Document(words=[contents], tags=[Doc_ID,Equip_ID]) * contents don't contain equip_id information
--

2.7 Doc2Vec 학습 결과 평가

본 연구는 doc2vec 학습 결과의 타당성을 평가하기 위해 적중률(Scoring)을 산출하였다. 학습되지 않고 장비 ID가 포함되지 않은 특정 무기체계 A의 고장 정비 문서를 벡터화 하여, 해당 벡터와 가장 유사한(Most Similar) 상위 장비 ID 1개 또는 3개를 각각 식별한다. 식별된 장비 ID에 무기체계 A가 포함되어 있는지를 확인하여, 이를 적중률(Scoring)로 계산하였다. 이를 Code 화하여 표현 하면 Table 8.과 같다.

Table 8. doc2vec Scoring

Identify Most Similar 1 Equipment (Most Similar topn=1)
<pre> ivec=model.infer_vector(doc_words=contents) infer_similar=model.docvecs.most_similar(positive=[ivec],topn=topn) for i in range(len(contents)): if Equip_ID in infer_similar : score+=1 else : pass print (score/len(contents)) *contents : not trained data </pre>
Identify Most Similar 3 Equipments (Most Similar topn=3)
<pre> ivec=model.infer_vector(doc_words=contents) infer_similar=model.docvecs.most_similar(positive=[ivec],topn=topn) for i in range(len(contents)): if Equip_ID in infer_similar : score+=1 else : pass print (score/len(contents)) *contents : not trained data </pre>

적중률 계산 시 상위 3개까지의 유사장비 ID를 식별한 이유는 유사한 기능을 하는 장비의 경우 고장 정비 내역 또한 유사하기 때문에 이를 고려한 범위 설정이다. 예를 들어 유사 규모의 레이더 장비 A, B가 있다면, 주요 기능인 RF 송수신 출력 및 레이더 구동에 대한 고장

정비 내역이 많이 있을 것으로 판단되며, 이는 고장 정비 내역을 기준으로 학습 시, 정확하게 두 장비를 식별 할 수 없음을 의미한다. 따라서 유사한 장비 군(Cluster)의 크기를 3개로 설정하여, 해당 군(Cluster)에 해당 장비 (A or B)가 존재하면 학습이 잘 되어 있는 것으로 판단 하였다. 적중률(Scoring) 확인 결과, 유사장비 ID 1가지를 식별한 경우의 적중률은 72%, 3가지를 식별한 경우는 87%로 나타났다(Table 9. 참조).

Table 9. Scoring Results

구분	Score
Identify Single Similar Equip ID (topn=1)	0.72
Identify 3 Similar Equip ID(topn=3)	0.87
* Test set size (Total Number of Docs to get score) : 6,382 Docs	

2.8 Doc2vec 학습 결과 활용

2.8.1 대표 고장 정비 유형 식별

대표 고장정비 유형은 각 장비 ID 벡터와 정비 문서 간 Cosine Similarity를 통해 식별 하였다. 예를 들어, 함상 레이더 장비(ID 14)는 정비문서 Doc_2106과 Cosine Similarity가 0.423으로 가장 높은 것으로 나타나 해당 문서에 포함된 정비내용이 대표 고장정비 유형인 것으로 나타났다(Table 10. 참조). 이는 유사 장비 개발 시, Doc_2106에 나타난 관련 기능 및 구성품에 대한 보다 신뢰성 있고 고장에 강건한 설계 방안을 적용해야 하며, 운영 유지 지원 방안 (보급품 설정, 예방정비, 보수정비, CSP 등)이 마련되어야 함을 의미한다.

Table 10. Representative Failure Mode for Equip_ID=14

	Equip_ID = 14	
	Doc ID	Cosine Similarity
1st	Doc_2106	0.423
2nd	Doc_2719	0.414
3th	Doc_2672	0.413
4th	Doc_51	0.406
5th	Doc_1540	0.403
6th	Doc_72	0.4
7th	Doc_2488	0.398
8th	Doc_2200	0.396
9th	Doc_2723	0.394
10th	Doc_2015	0.394

2.8.2 시각화(Visualization), 군집화(Clustering)

본 연구는 각 무기체계 ID 벡터 간 Cosine Similarity를 통해 무기체계 간 고장 정비 유형 간 유사도(Similarity)를 식별 하였다. 예를 들어, 함상 레이더 장비(ID 14)는 XXX 장비(ID 10)와 유사도가 0.259로 가장 높은 것으로 나타났다(Table 11 참조). 함상 레이더 장비(ID 14)와 유사한 장비 개발 시 유사도가 높은 장비들의 대표 고장 정비 내역을 검토하여 보다 가용도 높은 장비 개발에 활용될 수 있다.

Table 11. Most Similar Equipment

	Equip_ID = 14	
	Equip_ID	Cosine Similarity
1st	10	0.259
2nd	27	0.164
3th	15	0.163
4th	78	0.161
5th	63	0.160
6th	72	0.152
7th	45	0.150
8th	6	0.147
9th	43	0.146
10th	28	0.143

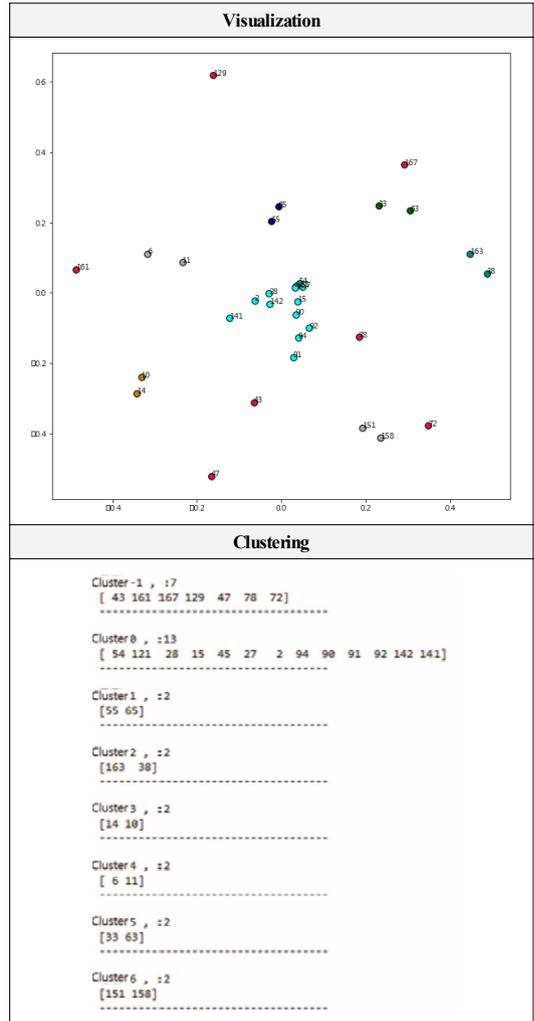
또한, 본 연구는 유사장비 식별 결과를 바탕으로 총 34 개의 무기체계에 대해 시각화(Visualization) 및 군집화(Clustering)를 수행하였다. 시각화(Visualization)와 군집화(Clustering)는 각각 t-sne (sklearn.manifold import TSNE)와 DBSCAN (sklearn.cluster import DBSCAN)를 적용하였다. 군집화 결과 34개의 무기체계(or 구성품)은 6개의 군집으로 Clustering 됨을 알 수 있다(Table 12 참조).

고장 정비 내역 기준 무기체계 전체에 대한 시각화(Visualization) 및 군집화(Clustering)는 개발 초기 유사 장비 파악을 하여, 주요 검토 대상 장비를 식별하는데 활용될 수 있다.

3. 결론

본 논문에서는 과거 자료로만 존재 했던 무기체계의 고장 정비 문서를 doc2vec을 적용하여 기계학습하는 방

Table 12. Weapon System(or Equipment) Visualization and Clustering



법을 제시하였으며, 학습된 결과를 적중률(score) 및 유사도(similarity)를 통해 대표 고장 정비 문서 및 고장유형을 식별하는 방법을 제시하였다. 또한, 전체 무기체계 간 유사성을 식별하기 용이하도록 시각화(Visualization)와 군집화(Clustering)를 수행하였다.

방대한 무기체계에 대한 고장 정비 데이터를 단 시간에 학습하여 무기 체계(or 장비) 별 대표 고장 정비 내역을 식별해 제시하는 방안은 새로운 무기체계 개발 시 매우 유용하게 활용될 수 있다. 유사장비에 대한 과거 고장 정비 내역은 현재 및 미래 개발 예정의 무기체계에 대하여 주요 강건 설계 대상 선정 기준으로 활용 되어 신뢰성 높은 설계를 적용 할 수 있으며, 주요 군수지원 대상

품목으로 식별하여, 군수지원 요소 개발 시 가용성을 확보 할 수 있는 방안을 마련 할 수 있기 때문이다.

그러나 유사성(Similarity), 시각화(Visualization), 군 집화(Clustering)는 비지도 학습(Unsupervised Learning)에 해당하므로 학습 결과로 제시된 내용에 대한 보수적인 평가가 필요하다[9]. 적중률(Scoring)로 정의된 학습 내용의 타당성 확인 방안을 확장하여, 학습 결과에 대한 관련 전문가의 평가도 필요하다.

References

- [1] Sunghoon Cho, Sukho Kang, "Machine Learning(AI) industrial application", *Industrial Engineering Magazine* 23(2), pp.34-38, 2016.06
- [2] Establishment of follow-up support system for export of weapons system, Konkuk Univ. Industry-Academia Collaboration Foundation Military Vision Lab, pp. 114-115, 2015.12
- [3] Sooyune Jeon, Donghun Lee, Manjae Bae, "Study on the Application Method of Munition's Quality Information based on Big Data", *Journal of the Korea Academia-Industrial cooperation Society*, Vol 17, pp.315-325, 2016
- [4] Hyun-jung Kim, "Big Data Concept and Big Data Analysis Technique", Seminar, Korean Transport Institute.
- [5] Establishment of follow-up support system for export of weapons system, Konkuk Univ. Industry-Academia Collaboration Foundation Military Vision Lab, p. summary-3, 2015.12
- [6] Jongmoon Rhee, Jongshin Lee, Seungryool Lee, Kyungduk Park, "A Study on FMECA based on failure rate and cost of occurrence", *Korean Institute Of Industrial Engineers*, pp.841-845, 2010.11
- [7] Andreas Muller, Sarah Guido(2017), Introduction to Machine Learning with Python, O'REILLY,
- [8] Gavagai, A BRIEF HISTORY OF WORD EMBEDDINGS (AND SOME CLARIFICATIONS), Gavagai, 2015.9.30., Available From: <https://www.gavagai.se/blog/2015/09/30/a-brief-history-of-word-embeddings> (accessed Mar., 30, 2018)
- [9] Mikolov Tomas, Sutskever Ilya, Chen Kai, Corrado Greg, Dean Jeffrey. "Distributed Representations of Words and Phrases and their Compositionality", In *Advances on Neural Information Processing System*, pp.3111-3119, 2013
- [10] Lau, Baldwin, An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation, arXiv.org, 2016.7.19.

박 연 경(Yun-Kyung Park)

[정회원]



- 2000년 3월 ~ 2006년 6월 : 고려대학교 공과대학 산업공학과 (산업공학학사)
- 2006년 7월 ~ 현재 : LIG넥스원 ILS 연구센터
- 2017년 3월 ~ 현재 : 고려대학교 컴퓨터정보통신대학원 소프트웨어공학과

<관심분야>

Reliability, Machine Learning, ILS

이 혜 원(Hye-Won Lee)

[정회원]



- 2008년 2월 ~ 2012년 2월 : 한국기술교육대학교 디자인공학과 (공학학사)
- 2012년 2월 ~ 2014년 2월: 포항공과대학교 산업경영공학과 (공학석사)
- 2014년 2월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

Human Factors, Reliability, ILS

김 상 문(Sang-Moon Kim)

[정회원]



- 1990년 3월 ~ 1995 2월 : 성균관대학교 공과대학 산업공학과 (산업공학학사)
- 1995년 6월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

Reliability, Logistic Support, Availability, ILS