

상용 부품 비정형 데이터와 인공 신경망을 이용한 부품 단종 예측 방안 연구

박연경*, 이익도, 이강택, 김두정
LIG빅스원 ILS연구소

Study on predicting the commercial parts discontinuance using unstructured data and artificial neural network

Yun-kyung Park*, Ik-Do Lee, Kang-Taek Lee, Du-Jeoung Kim
ILS R&d IAB, LIG Nex1 Co., Ltd

요약 기술의 발전으로 다양한 부품의 개발 및 상용화는 가능 하였으나, 이에 따라 부품의 단종 주기는 단축 되었다. 이는 수천 품목 이상의 부품을 활용하여 개발하고, 장기간 운영하는 무기체계의 수리 부속 보급을 어렵게 하였으며, 무기 체계 운용 가용도 저하의 주요 원인으로 작용하였다. 이러한 문제를 해결하기 위하여 미국 등은 전담 기구를 만들어 대응하고 있으며, 국내에서는 상용 부품단종 예측도구를 활용하여 단종을 예측하고 관리하고 있다. 하지만 상용 부품단종 예측도구에서 단종 정보가 제시되지 않는 부품에 대한 대응 및 관리는 부재한 실정이다. 이에 본 연구에서는 상용 부품단종 예측도구에서 제공하는 부품에 대한 정형, 비정형 빅데이터를 수집하고, 데이터 전처리 및 Embedding 과정을 거쳐, 신경망 학습 알고리즘을 적용하여, 상용 부품에 대한 단종 정보 (LC Risk, YTEOL)를 예측하는 방안을 제시하였다. 또한 제시된 모델의 예측 성능을 데이터 기술 통계량과 비교 평가 하여, 본 연구에서 제시한 학습 모델의 타당성을 검증 하였다. 결론에는 본 연구의 활용 방안과 한계점 및 발전 방향에 대하여 기술 하였다.

Abstract Advances in technology have allowed the development and commercialization of various parts; however this has shortened the discontinuation cycle of the components. This means that repair and logistic support of weapon system which is applied to thousands of part components and operated over the long-term is difficult, which is the one of main causes of the decrease in the availability of weapon system. To improve this problem, the United States has created a special organization for this problem, whereas in Korea, commercial tools are used to predict and manage DMSMS. However, there is rarely a method to predict life cycle of parts that are not presented DMSMS information at the commercial tools. In this study, the structured and unstructured data of parts of a commercial tool were gathered, preprocessed, and embedded using neural network algorithm. Then, a method is suggested to predict the life cycle risk (LC Risk) and year to end of life (YTEOL). In addition, to validate the prediction performance of LC Risk and YTEOL, the prediction value is compared with descriptive statistics

Keywords : Machine Learning, Big Data, Fasttext, Neural Network, DMSMS, LC Risk, YTEOL

1. 연구 배경

과학과 기술이 빠르게 발전하면서 제품을 사용하는 사

용자는 다양한 환경에서 정밀하게 기능하는 장비를 원하고 있다. 이것은 일상생활에서 접하는 전자제품 뿐만 아니라 무기체계에도 적용되는 기술 발전의 추세이다. 반면

*Corresponding Author : Yun-Kyung Park(LIG Nex1 Co., Ltd)
email: ykpark2002@lignex1.com

Received July 10, 2019

Accepted October 4, 2019

Revised August 5, 2019

Published October 31, 2019

이러한 기술 발전 속도는 부품류의 빠른 교체주기와 짧은 수명주기를 갖도록 하고 있다[1-2].

우리 군은 현재 인공지능 등의 4차 산업혁명 핵심 요소를 근간으로 하여 현대 전장을 스마트 전장으로 갖추고 있으며 무기체계에 고도의 정밀화와 자동화를 갖출 것을 요구하고 있다. 이러한 요구를 달성하기 위해 대부분의 무기체계는 수천에서 수만 개의 최신 부품을 활용하여 약 5~10년간 개발한다.

개발된 무기체계는 작전지역에 배치되어 25~30년 이상 사용하고 있다. 최근 전자부품의 수명주기가 4~7년 정도로 예상되므로, 무기체계가 작전지역에 배치되는 순간부터 적용된 부품의 진부화는 이미 시작된다고 볼 수 있다[3-4].

진부화된 부품은 결과적으로 부품 단종(DMSMS, Diminishing Manufacturing Sources and Material Shortage) 문제를 발생시키며 무기체계의 정비 지연을 초래하여 전투력 유지에 영향을 미치게 된다[5].

미국을 비롯한 선진국에서는 무기체계 수명주기 기간 중 부품 단종으로 인해 많은 어려움을 겪었으며, 이로 인한 문제점을 최소화하기 위하여 여러 가지 기술 개발과 함께 관련 규정 정비 및 전담기구 운영 등 많은 노력을 기울이고 있다[6].

대한민국 군에서도 무기체계에 발생하는 부품 단종 문제의 중요성을 인식하고 있다. 이에 따라 국방부에서 부품단종관리 업무지시(국지시 제15-023호, 2015.9.1.)를 발령하여 총수명주기체계관리(TLCSM) 관점의 최적화된 부품단종관리 체계를 구축하고 있다. 또한 방위사업청에서는 관련 부서를 만들어 대응 하는 등 정책적인 대처방안과 부품단종예측 도구를 활용할 것을 제시하고 있다[7].

위와 같은 대응책에 따라 국내 방위산업체에서는 부품 단종 예측도구를 활용하여 상용 부품에 대한 단종을 예측하고 있다. 하지만 예측도구가 다수이며, 각 도구마다 제공하는 단종 정보와 부품 정보의 범위가 상이하다. Park et al(2018)이 연구한 바에 따르면 특정 무기체계의 BOM에 대하여 각 단종 예측도구는 63.5%, 43.5%, 55.3%의 검색률을 보이는 것으로 조사되었다[8].

예측 도구에서 단종 정보를 제공하지 않는 부품에 대한 수명주기 예측 방안으로는 Scorecard Approach, Availability Factor Method 등을 활용 할 수 있다. Solomon et al 은 두 방법에 대한 요약과 한계를 기술 하며, 부품에 대한 Sales Data 등을 이용한 방법을 제안 하기도 하였다[9]. 또한 위와 같은 방법론에 Peter et al(2007)은 데이터 마이닝 기법을 더해 부품 단종 예측 결과의 정확성을 향상시키기도 하였다[10].

그러나 소요군 또는 방위산업체 등 완성품을 운용 및 제조하는 조직에서 기 제시된 방법론을 적용하여, 부품 단위의 수명 주기 예측을 적용하기에는 필요한 제반 데이터 접근 및 획득에 한계가 있으므로, 단종예측 도구에서 제공되지 않는 부품에 대한 수명 주기 예측은 현실적으로 제한된다.

모든 부품에 대한 단종 정보는 무기체계 개발부터 폐기까지 관리하여 운용 간 문제가 발생하지 않도록 준비 하는 것이 이상적이며, 부품에 대한 수명주기를 예측하여 수리 부속 보급 및 장비 운용 계획을 세우는 것이 중요하다.

앞서 밝힌 바와 같이 부품 단종에 대한 연구는 꾸준히 이루어지고 있으나, 최근 대두되고 있는 기계학습 등의 인공지능 기술을 부품 단종 대응을 위해 적용한 사례는 거의 없다.

따라서 본 연구에서는 기계 학습 방법을 통해 상용 단종 예측도구에서 제공하는 부품정보와 단종정보를 학습 하여, 예측하는 방안을 제시한다. 또한, 예측한 값을 검증 데이터와 비교하여, 본 연구에서 제안한 방안의 단종 정보 예측 성능을 평가 하였다.

2. 본론

2.1 연구 프로세스

본 논문의 연구 프로세스는 Fig. 1과 같다. 1 단계에서는 상용 부품 정보제공 도구 (Web 기반 정보 제공 도구)를 활용하여, 부품에 대한 비정형 데이터 및 Life Cycle Risk(이하 LC risk), Year To End Of Life (이하 YTEOL) 정보 54,665 건을 수집 하였다. 정보 수집은 자체 개발한 Web Crawler를 활용 하였다. 2단계에서는 수집된 부품 비정형 데이터의 Embedding을 수행하여, 기계학습 입력에 필요한 각 부품 데이터의 벡터화를 수행하였으며, 3단계에서는 기계학습을 위한 학습 모델로 Fully connected Neural Net을 제안하고, 입력으로 Embedding Value를, 출력으로는 LC Risk 또는 YTEOL을 학습하는 모델을 제시 하였다. 4단계에서는 학습된 모델을 이용하여, 비 학습 품목에 대한 예측을 하고 이에 대한 성능을 검증하였다. 총 54,665 개의 부품 데이터 중 75%인 40,998 개의 데이터를 학습에 활용 하였으며, 25%인 13,667 개의 데이터를 검증에 활용 하였다. 또한 자료의 불균형에 의한 성능 왜곡 여부를 확인하기 위하여, LC Riks와 YTEOL의 기술통계량을 기준으로 학습 여부를 확인 하였다.

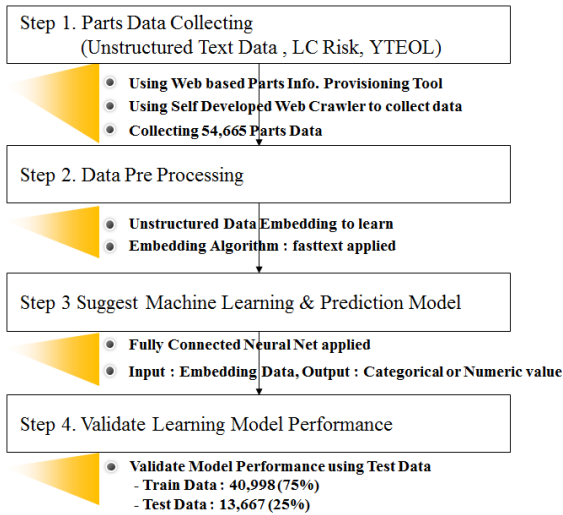


Fig. 1. Research process

2.2 데이터 수집

본 연구에서는 모 방산업체 수명주기관리 시스템에 등록된 상용 부품 목록을 기준으로, 상용 부품 정보제공 도구를 활용하여, 해당 부품에 대한 비정형 데이터 및 LC Risk , YTEOL 정보 54,665 건을 수집 하였다.

다만 YTEOL의 경우 단종 정보 예측 일자 기준 가용 연도를 표기하고 있어, 모든 부품에 대하여 데이터를 수집한 일자기준('19.04.13)으로 보정하였다. 예를 들어, '17.11.22에 예측된 YTEOL이 8.3년이라면 Data 수집 일자 기준('19-04-13)으로 보정하여 6.9년을 적용하였다.

수집한 부품 데이터는 Table 1과 같이 각 상용 부품에 대한 부품번호, 생산자명, Category, 부품 설명, 생산자가 제공한 부품 설명 상세, 생산자가 제공한 Package 설명 상세, 수명주기 위험도 분류, 예측된 가용 연도 등을 포함하고 있다.

Table 1. Data Collection Details

| Category | Details |
|------------------|---|
| Data Type | Unstructured Text Data, Numeric Data |
| Num of Part Data | 54,665 |
| Contents | <ul style="list-style-type: none"> Part Number Mfr Name Category Description Manufacturer Part Description Manufacturer Package Description LC Risk Estimated YTEOL |

수집된 상용 부품 정보 예시는 Table 2와 같다. 부품에 대한 특징 및 제원이 비정형 데이터로 기술되어 있으며, 단종에 대한 위험 수준은 High, Med, Low 3가지의 Categorical 데이터로 표현되어 있다. 또한 단종 예측시기(YTEOL)는 Numeric 데이터로 제시되어 있다.

Table 2. Collected Data Example

| | Details |
|----------------------------------|--|
| Part Number | 0402B561K500CT |
| Mfr Name | Walsin |
| Category | CAPACITOR,CHIP CERAMIC 1005 |
| Description | Ceramic Capacitor, Multilayer, Ceramic, 50V, 10% +Tol, 10% -Tol, X7R, 15% TC, 0.00056uF, Surface Mount, 0402 |
| Manufacturer Part Description | MULTILAYER CERAMIC CAPACITOR |
| Manufacturer Package Description | CHIP |
| LC Risk | LOW |
| Estimated YTEOL | 7.82 |

2.3 데이터 전처리(Pre-Processing)

본 연구에서는 비정형 데이터를 신경망을 이용하여 학습을 수행하기 위하여, fasttext 알고리즘을 이용한 Embedding을 적용하였다. Embedding 구현 및 Hyper Parameter는 Table 3, Table 4와 같다.

각 부품의 비정형 데이터는 fasttext 알고리즘으로 벡터화되어 신경망 입력값으로 활용하였다.

Table 3. Library and Implementation

| | Description |
|----------|-----------------|
| Language | Python 3.7.1 |
| Library | gensim.fasttext |

Table 4. Fasttext Hyper Parameter

| Param | Description |
|-----------|--|
| size | Dimensionality of the feature vectors |
| window | The maximum distance between the current and predicted word within a sentence. |
| min_count | Ignores all words with total frequency lower than this. |
| epochs | Number of iterations (epochs) over the corpus |
| negative | If > 0, negative sampling will be used, the int for negative specifies how many "noise words" should be drawn (usually between 5-20). If set to 0, no negative sampling is used. |

| | |
|--------------------|--|
| alpha | The initial learning rate. |
| min_alpha | Learning rate will linearly drop to min_alpha as training progresses |
| hs | If 1, hierarchical softmax will be used for model training. If set to 0, and negative is non-zero, negative sampling will be used. |
| sg | Training algorithm: skip-gram if sg=1, otherwise CBOW |
| word_ngrams | If 1, uses enriches word vectors with subword(n-grams) information. If 0, this is equivalent to Word2Vec. |
| min_n | Minimum length of char n-grams to be used for training word representations. |
| max_n | Max length of char ngrams to be used for training word representations. Set max_n to be lesser than min_n to avoid char ngrams being used. |

2.4 학습모델 제안 및 학습

본 연구에서는 학습 모델을 구현하기 위하여 Table 5와 같은 Language 및 Library를 사용하였다.

Table 5. Library and Implementation

| | Description |
|-----------------|--------------|
| Language | Python 3.7.1 |
| Library | keras.layers |

본 연구에서 제안하는 학습모델과 입출력 데이터 및 신경망에 대한 설명은 Fig. 2와 같다.

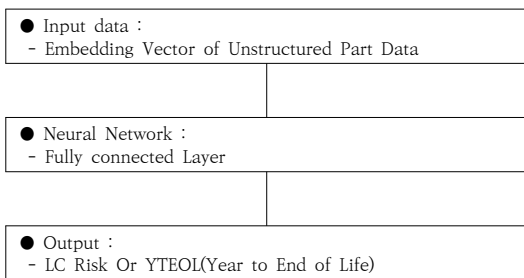


Fig. 2. Learning Model for DMSMS Prediction.

학습 데이터는 총 54,665 건의 데이터 중 75%인 40,998 건의 데이터를 무작위로 선정하여 수행하였으며, 25%인 13,667 건의 데이터는 학습 모델 성능의 검증용으로 활용 하였다.

2.5 모델 검증

2.5.1 LC Risk와 YTEOL 기술 통계량

데이터 셋의 자료 불균형 문제로 인한 학습 모델의 성능을 적절하게 평가 할 수 없는 경우가 발생 할 수 있다. 가령 부품 데이터 90%이상의 LC Risk가 'Med'라면, 아무 학습을 거치지 않은 모델에서도 90% 이상의 정확도 성능을 낼 수 있다. 학습된 모델이 데이터 셋 불균형에 의한 성능을 초과하여, 적절하게 학습되었는지 확인하기 위해서는 성능에 대한 적절한 기준을 정하는 것이 필요하다. 본 연구에서는 데이터 셋의 LC Risk 및 YTEOL에 대한 기술 통계량을 산출하여, 모델의 성능의 평가 기준으로 선정하였다. Excel S/W의 데이터 분석 기능을 이용하여 LC Risk 및 YTEOL에 대한 기술 통계량을 계산 하였으며, 그 결과는 Table 6, Table 7과 같다.

단, YTEOL에 대한 결측치 Data가 존재 할 경우 해당 부품 LC Risk의 평균 YTEOL을 적용 하였다. 즉, 부품 '0402B474K100CT' 의 LC Risk가 'Low' 이나, YTEOL 값이 결측되어 있다면, 'Low'의 평균 YTEOL인 7.0 년을 적용 하였다.

Table 6. LC Risk Descriptive Statistics

| | Count | Ratio |
|-----------------------|--------|-------|
| Total Data | 54,665 | 100% |
| LC Risk 'Low' | 23,089 | 42.2 |
| LC Risk 'Med' | 16,580 | 30.3 |
| LC Risk 'High' | 14,996 | 27.4 |

Table 7. YTEOL Descriptive Statistics

| | Count |
|-------------------------------|--------|
| Total Data | 54,665 |
| Mean | 3.06 |
| Std | 3.51 |
| Confidence Level (95%) | ±0.03 |

학습 모델의 LC Risk의 식별 성능은 42.2%를 초과해야 하며, 예측된 YTEOL과 Real Value 간의 표준편차 (Std)는 3.51 미만이어야 데이터 셋의 불균형에 따른 성능이 아닌 학습 모델의 적절성에 의한 성능으로 판단할 수 있다.

2.5.2 LC Risk 학습, 예측, 검증

본 논문 2.4 에서 제시한 학습 모델을 이용하여, 40,998 개의 상용 품목의 LC Risk를 학습 하였으며, 13,667 개의 상용 품목에 대하여 LC Risk를 예측하여 성능을 검증 하였다. 학습 모델은 Fully connected Neural Net을 적용하였으며, 6 layer, 256 unit/layer를 적용하였다. Categorical Data를 Output으로 학습하는 모델이므로, Output value 적용 함수로는 softmax를 설정하였다. Epochs는 Heuristic을 적용하여 100 회로 설정하였다.

Hyper Parameter 는 Table 8과 같으며, LC Risk 예측에 대한 성능은 아래 Fig. 3과 같다.

Table 8. Learning Model Hyper Parameter

| Neural Network | Arguments |
|----------------|--------------------------------|
| Dense Layer | 6 layers |
| | 256 units/layer |
| | relu |
| | 0.2 |
| Output | softmax |
| optimizer | Adam |
| loss function | categorical_crossentropy |
| batch_size | 32 |
| epochs | 100 |
| Input Value | Embedding Value |
| Output Value | LC Risk (Categorical value) |



Fig. 3. Training and Test Accuracy.

학습 모델에 대한 Validation Accuracy는 0.85 이며, Precision, Recall, f1 Score는 Table 9와 같다.

Table 9. Prediction Performance of Learning Model

| | Precision | Recall | f1-score |
|--------------|-----------|--------|----------|
| High | 0.79 | 0.8 | 0.8 |
| Low | 0.89 | 0.88 | 0.88 |
| Med | 0.85 | 0.85 | 0.85 |
| micro avg | 0.85 | 0.85 | 0.85 |
| macro avg | 0.84 | 0.84 | 0.84 |
| weighted_avg | 0.85 | 0.85 | 0.85 |

2.5.3 YTEOL 학습, 예측, 검증

본 논문 2.4 에서 제시한 학습 모델을 이용하여, 40,998 개의 상용 품목의 YTEOL을 학습 하였으며, 13,667 개의 상용 품목에 대하여 YTEOL을 예측하여 성능을 검증 하였다. 학습 모델은 Fully connected Neural Net을 적용하였으며, 6 layer, 256 unit/layer를 적용하였다. Epochs는 Heuristic을 적용하여 200 회로 설정하였다.

Hyper Parameter 는 Table 10과 같으며, LC Risk 예측에 대한 성능은 Fig. 4와 같다.

Table 10. Learning Model Hyper Parameter

| Neural Network | Arguments |
|----------------|----------------------------|
| Dense Layer | 6 layers |
| | 256 units/layer |
| | relu |
| | 0.3 |
| optimizer | Adam |
| loss function | mean square error |
| batch_size | 32 |
| epochs | 200 |
| Input Value | Embedding Value |
| Output Value | YTEOL (Numerical value) |

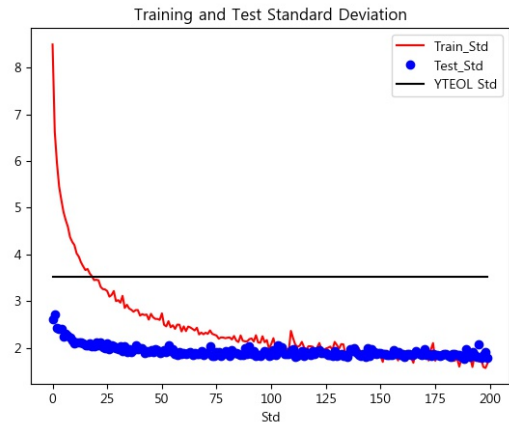


Fig. 4. Training and Test Accuracy.

2.5.3 학습 결과 종합

본 연구에서 제안한 모델을 이용하여, LC Risk와 YTEOL을 학습하고 예측한 결과는 Table 11.과 같다. LC Risk의 예측 성능은 85% 이며, YTEOL의 예측 표준 편차는 1.77 으로, 본 연구에서 학습 효과 여부에 대한 기준치로 제시한 기술 통계량 42%와 3.51 보다 좋은 예측 성능 보임을 알 수 있다. 즉, 본 모델을 이용한 학습에 대한 효과가 있음을 알 수 있다.

Table 11. Prediction Performance of Learning Model

| | Descriptive Statistics | Model Performance |
|---------|------------------------|--------------------------------|
| LC Risk | 0.42 ('Low' Ratio) | 0.85 (Weighted_avg) |
| YTEOL | 3.51 (YTEOL Std.) | 1.77 (Std. of Real-Prediction) |

3. 결론

본 연구에서는 단종 정보를 수집 할 수 있는 부품에 대한 비정형 데이터를 신경망을 이용하여 학습하는 모델을 제안하였으며, 이를 이용하여 부품 단종 정보가 없는 품목에 대한 단종 정보 (LC Risk/YTEOL)를 예측하였다. 또한 예측된 단종 정보(LC Risk/YTEOL)의 성능을 측정하였고, 상식 수준 기준점을 설정하여, 예측 결과가 무작위에 의한 결과가 아닌, 학습된 모델에 의한 성능임을 입증하였다.

모든 부품에 대한 단종 정보는 식별하여, 관리/ 대응 하는 것이 이상적이거나, 단종 정보를 식별하기 어려운 현실적인 요인으로 인해 단종 정보 식별이 제한되는 부품에 대하여 본 연구가 활용되어 짐을 기대 한다.

단, 본 연구에 활용된 부품데이터는 IC, Capacitor, Resistor, Inductor 등 한정적인 부품 Category 내에서 학습되고 예측되어진 결과이므로, 보다 다양한 Category 및 많은 데이터를 수집하여 Category 별 단종 특성을 연구/예측/평가 할 필요가 있다. 또한 학습 모델 적용을 RNN, CNN, LSTM 등 다양화 하여 시도함으로써, 부품 단종 예측에 최적화된 모델을 연구 할 필요도 있다.

Manufacturing Source and Material Shortages Management Cost Analysis to Select Optimization Alternatives", *The Institute of Electronics and Information Engineers*, Korea, pp. 226-228, Nov. 2018.

- [2] Won-Cheol Baek, *A Study on Discontinue Parts Management Considering the Korean Military Arms Properties*, Master's thesis, Hansei University, 2014.
- [3] Seok-Gon Kim, Jong-Jun Yoon et al, *World of Naval Missile System*, Hantee media, pp. 25-30. 2016
- [4] Won-Jong Chun et al. "A Study on DMSMS Management for Guided Missiles at Operation and Sustainment Stage", *Journal of Korea Society for Naval Science and Technology*, Vol. 1, pp. 7-13, 2018.
- [5] Gi-Hong, Kim, "A Study on Parts Obsolescence Prevention and Management Plan", *Defense & Technology*(317), pp. 56-65, 2005.
- [6] Kwang-hyo Park, Bo-hyun Shim, "A study on the Implementation and Development of the systematic DMSMS Management", *Journal of The Institute of Electronics and Information Engineers* Vol.56, NO.2, 2019
DOI : <https://dx.doi.org/10.5573/ieie.2019.56.2.33>
- [7] Defense Agency for Technology and Quality, *The Guidelines on DMSMS Management*, Defense Agency for Technology and Quality, pp. 1-9. 2017
- [8] Kwang-hyo Park, Bo-hyun Shim, "A study on the Research and Application for the Private DMSMS prediction tool", *Journal of The Institute of Electronics and Information Engineers* Vol.55, NO.12, Dec. 2018
DOI : <https://dx.doi.org/10.5573/ieie.2018.55.12.35>
- [9] Rajeev Solomon et al., " Electronic Part Life Cycle Concepts and Obsolescence Forecasting", *IEEE Trans. on Components and Packaging Technologies*, pp. 707-717, Dec. 2000.
DOI : <https://doi.org/10.1109/6144.888857>
- [10] Peter Sandborn et al., "A Data Mining Based Approach to Electronic Part Obsolescence Froecasting", *IEEE Trans. on Components and Packaging Technologies*, pp. 397-401, Sept. 2007.
DOI : <https://doi.org/10.1109/TCAPT.2007.900058>

References

- [1] Kwang-Hyo Park et al., "A Study On the Diminishing

박 연 경(Yun-Kyung Park)

[정회원]



- 2006년 8월 : 고려대학교 산업공학과 (공학학사)
- 2019년 8월 : 고려대학교 소프트웨어 공학과 (공학석사)
- 2006년 7월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

Machine Learning, Logistic Support, ILS

김 두 정(Du-Jeoung Kim)

[정회원]



- 2012년 2월 : 성균관대학교 시스템경영공학과 (공학학사)
- 2014년 2월 : 성균관대학교 산업공학과 (공학석사)
- 2014년 1월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

HCI, UX, Accessibility, Ergonomics

이 익 도(Ik-Do Lee)

[정회원]



- 2001년 3월 : 성균관대학교 산업공학과 (공학학사)
- 2009년 8월 : 성균관대학교 기술경영학과 (공학석사)
- 2001년 1월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

DMSMS, Logistic Support, ILS

이 강 택(Kang-Taek Lee)

[정회원]



- 2012년 2월 : 성균관대학교 시스템경영공학과 (공학학사)
- 2014년 2월 : 성균관대학교 기술경영학과 (공학석사)
- 2015년 1월 ~ 현재 : LIG넥스원 ILS 연구센터

<관심분야>

기술경영, 기술사업화, ILS