

Received: 2026/05/27
Revised: 2026/06/08
Accepted: 2026/06/21
Published: 2026/06/30

***Corresponding Author:**

Youngsoo Namgoong

20, Daewangpangyo-ro 851beon-gil, Sujeong-gu,
Seongnam-si, Gyeonggi-do, 13449, Republic of Korea
E-mail: youngsoo.namgoong@ligdna.com

Abstract

본 논문에서는 무기체계 신뢰도 분석에 활용되는 유사장비비교법의 정량화를 위해 다양한 데이터 항목별 가중합 방식을 도입한 통합 유사도 산출 방법론을 제안한다. 제안 방법론을 대표성있는 데이터에 적용한 결과, 유사도가 높은 장비를 효과적으로 식별함으로써 재현성 있고 객관적인 유사 장비 선정이 가능하며 대규모 데이터에 대한 유사도 계산 자동화로 분석 효율성을 크게 향상시킬 수 있음을 확인하였다.

In this paper, we propose an integrated similarity calculation methodology that incorporates various weighted-sum approaches for different data items to quantify the similar-item prediction method used in weapon-system reliability analysis. Applying the proposed methodology to representative data demonstrated that it can effectively identify equipment with high similarity, enabling reproducible and objective selection of comparable systems. Additionally, automating similarity calculations for large datasets significantly improves analysis efficiency.

Keywords

유사장비비교법(Similar-Item Prediction Method), 유클리디안 거리(Euclidean Distance), 맨해튼 거리(Manhattan Distance), 자카드 유사도(Jaccard Similarity)

유사장비 데이터 기반 신뢰도 예측 방안 연구

Similar-Item Prediction Method: A Data-Driven Approach Based on Similarity and Distance

남궁영수^{1*}, 석준영¹, 손은호², 박병호²

¹LIG D&A 미래전장IPS연구소 연구원

²LIG D&A 미래전장IPS연구소 선임연구원

Youngsoo Namgoong^{1*}, Junyeong Seok¹, Eunho Son², Byeongho Park²

¹Research engineer, Future Warfare Integrated Product Support R&D Lab., LIG Defense&Aerospace

²Research engineer, Future Warfare Integrated Product Support R&D Lab., LIG Defense&Aerospace

1. 서론

1.1 연구의 배경 및 필요성

현대의 무기체계는 첨단 기술의 집약체로서 그 구조가 고도화되고 복잡해짐에 따라 운용 유지의 효율성과 작전 지속 능력을 보장하는 RAM의 중요성이 날로 커지고 있다. 특히 국방 획득 절차에 있어 사업 초기 단계의 RAM 업무는 향후 발생할 총수명주기비용을 절감하고 설계 결함을 사전에 차단하기 위한 필수적인 절차로 제도화되고 있다[1,2].

RAM 업무 중 신뢰도 예측을 위한 대표적인 기법으로는 부품수량분석과 부품부하분석이 있으나, 상세 설계 정보와 부품 수준의 고장률 데이터가 요구되어 개발 초기 단계에서는 적용이 제한된다. 대안으로 기존에 개발 및 운용된 장비의 신뢰도 데이터를 활용하는 유사장비비교법이 실무적으로 널리 활용되고 있으나, 그 핵심 과정인 유사성 판단에 대한 정량적 기준이 제시되지 않음에 따라 분석자의 경험과 직관에 의존하고, 분석 결과의 재현성이 저하되는 문제가 발생한다[3,4].

이를 해결하기 위하여 설문조사를 반영하거나 AHP기법을 활용하여 분석 과정을 객관화하려는 시도가 있었으나 분석자 주관에 개입하는 것을 차단할 수는 없으며, 과거 개발 사례 데이터가 방대해짐에 따라 축적되는 수많은 데이터를 수동으로 탐색하는 비효율

성의 문제가 있다[5]. 품목명과 BOM구성의 유사도를 기준으로 정량 분석하려는 시도가 있었으나 구성품의 기능, 제원 등 구체적 정보가 고려되지 못하는 한계가 있다[6].

1.2 연구의 목적 및 내용

정보공학 및 데이터 마이닝 분야에서는 유클리디안 거리, 맨해튼 거리, 레벤슈타인 거리, 코사인 유사도, 자카드 유사도 등 여러 수학적 지표를 활용한 유사도 산출 방법이 활발히 연구되어 그 효과성이 검증된 바 있다. 하지만 이러한 기법들은 주로 추천 시스템이나 일반 텍스트 분석에 국한되어 왔으며, 국방 신뢰도 분석의 특수성을 반영한 장비 유사성 판단 도구로 활용된 사례는 찾을 수 없다.

본 연구는 기연구된 여러 유사도 산출 기법을 무기체계 신뢰도 분석의 유사장비비교법에 접목하고, 가중치를 두어 통합한 지수를 고안하여 그 효과를 판단한다. 이를 통해 유사장비 선정 과정에서 분석자의 주관성을 배제하고, 대량 데이터에 대한 유사도 추출을 자동화하여 효율성을 높이는 방법을 제안하고자 한다. 인공지능 및 데이터 기반 의사결정이 강조되는 추세에 맞춰, 다수의 기개발 사업 데이터 속에서 최적의 유사 품목을 과학적으로 탐색할 수 있는 사례로 제시한다.

2. 유사도 측정 기법

유사도(similarity)는 두 객체 간의 유사성을 정량적으로 표현하며, 수학적으로는 동일한 표현 공간에 속한 두 대상 $x, y \in X$ 에 대해 정의되는 실수값 함수 $\text{sim}: X \times X \rightarrow \mathbb{R}$ 로 나타낼 수 있다. 이는 두 대상이 공유하는 특성, 구조, 또는 의미적 관계의 정도를 하나의 척도로 환원하는 과정으로 이해할 수 있으며, 데이터의 표현 방식에 따라 벡터, 집합, 확률 분포 등 다양한 형태에 대해 정의될 수 있다. 유사도는 반드시 특정 범위에 제한될 필요는 없으며, 문제의 목적과 해석 방식에 $[0,1]$, $[-1,1]$, 혹은 비제한적인 실수 범위를 가질 수 있다. 따라서 유사도의 절대적 값 자체보다는, 서로 다른 대상 쌍 간의 상대적 크기를 기반으로 유사성의 정도를 비교하는 데에 의미가 있다. 또한 유사도는 거리 함수와 밀접하게 관련되며, 일반적으로 거리

와 단조 감소 관계를 이루는 변환을 통해 상호 변환이 가능하다. 이러한 특성으로 인해 데이터 간의 관계를 정량적으로 표현할 수 있으며, 주어진 데이터 포인트와 유사한 객체를 탐색하는 데 활용될 수 있다.

거리(distance)는 두 객체 간의 차이를 정량적으로 나타내는 값으로, 일반적으로 값이 작을수록 두 객체가 더 유사한 것으로 해석된다. 이러한 특성으로 인해 유사한 객체를 탐색하는 문제에서는 유사도 또는 거리 중 하나를 선택하여 사용할 수 있으며, 두 방식은 문제의 정의와 해석 방식에 따라 상호 대체하여 활용될 수 있다.

실제 응용에서는 데이터의 표현 방식과 분석 목적에 따라 다양한 유사도 및 거리 기반 방법이 활용된다. 이러한 방법들은 데이터의 형태와 표현 방식에 따라 서로 다른 방식으로 유사성을 정의하며, 그에 따라 계산 방식과 해석에서도 차이를 보인다. 따라서 특정 방법을 일괄적으로 적용하기보다는, 문제의 목적과 데이터의 특성을 고려하여 적절한 방법을 선택하는 것이 필요하다.

2.1 유클리디안 거리

유클리디안 거리(Euclidean distance)는 n 차원 유클리드 공간 상에서 두 점 사이의 최단 거리를 계산하는 방법으로, 가장 기본적인 거리 측정 중 하나이다. n 차원 공간에 속한 두 벡터 $x = (x_1, x_2, \dots, x_n)$ 와 $y = (y_1, y_2, \dots, y_n)$ 에 대해 유클리디안 거리는 각 차원에서 차이를 제곱하여 합한 뒤 제곱근을 취하는 방식으로 정의된다. 이는 다음과 같이 표현할 수 있다.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

데이터가 연속적인 수치형 벡터로 표현되는 경우 유클리디안 거리를 통해 직관적인 해석이 가능하다. 식 (1)에서 각 차원의 차이는 제곱을 통해 결합되므로, 값의 차이가 큰 차원일수록 전체 거리 계산에 더 큰 영향을 미치게 된다. 또한 유클리디안 거리는 좌표계의 방향이나 축의 설정에 영향을 받지 않으며, 공간을 회전시키더라도 두 점 사이의 거리 값은 변하지 않는다. 이러한 특성은 데이터의 표현 방식에 관계없이 두 객체 간의 상대적인 차이를 일관되게 반영할 수 있음을 의미한다.

유클리디안 거리는 모든 차원의 차이를 동시에 고려하여 두 객체 간의 전반적인 차이를 하나의 값으로 반영한다. 따라서 특정 속성의 일치 여부보다는 전체적인 값의 차이를 기반으로 유사성을 판단하는 데 적합하다. 그러나 차원의 수가 증가함에 따라 데이터 간 거리 차이가 상대적으로 감소하는 경향이 나타날 수 있으며, 이로 인해 거리 기반 비교의 구분력이 저하될 수 있다[7]. 따라서 고차원 데이터에 대해서는 거리의 해석에 주의가 필요하다.

2.2 맨해튼 거리

맨해튼 거리(Manhattan distance)는 n 차원 공간에서 두 점 사이의 거리를 각 차원에서의 절댓값 차이의 합으로 계산하는 방법이다. n 차원 공간에 속한 두 벡터 $x = (x_1, x_2, \dots, x_n)$ 와 $y = (y_1, y_2, \dots, y_n)$ 에 대해 맨해튼 거리는 다음과 같이 정의된다.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

식 (2)는 각 차원의 차이를 절댓값 형태로 누적하여 계산하며, 이는 좌표축 방향으로만 이동하는 경로의 길이를 측정하는 방식으로 해석할 수 있다. 맨해튼 거리는 각 차원의 차이를 선형적으로 반영한다는 점에서, 전체적인 차이를 반영하되 특정 차원의 값이 상대적으로 과도하게 강조되지 않는다는 특징을 가진다.

2.3 레벤슈타인 거리

레벤슈타인 거리(Levenshtein distance)는 두 문자열 간의 차이를 정량적으로 측정하기 위한 거리 중 하나로, 한 문자열을 다른 문자열로 변환하는 데 필요한 최소 편집 연산의 횟수로 정의된다. 여기서 편집 연산은 일반적으로 문자 삽입, 삭제, 치환을 포함한다.

두 문자열 s 와 t 에 대해 레벤슈타인 거리를 $d(s, t)$ 로 정의하면, 이는 다음과 같은 점화식을 통해 표현될 수 있다.

$$d(i, j) = \begin{cases} i & \text{if } j=0 \\ j & \text{if } i=0 \\ \min \left\{ \begin{array}{l} d(i-1, j) + 1 \\ d(i, j-1) + 1 \\ d(i-1, j-1) + \delta(s_i, t_j) \end{array} \right\} & \text{otherwise} \end{cases} \quad (3)$$

식 (3)에서 $d(i, j)$ 는 문자열 s 의 처음 i 개의 문자와 문자열 t 의 처음 j 개의 문자 사이의 거리를 의미하며, s_i 와 t_j 는 각각 문자열 s 와 t 의 i 번째와 j 번째 문자를 나타낸다. 또한 $\delta(s_i, t_j)$ 는 두 문자가 서로 같을 경우 0, 다를 경우 1의 값을 갖는 함수이다.

레벤슈타인 거리는 문자열의 길이와 문자 순서를 모두 고려하여 차이를 측정한다는 특징을 가지며, 단순한 문자 일치 여부 뿐 아니라 문자 간의 편집 과정을 통해 발생하는 구조적 차이를 반영할 수 있다. 그러나 문자열의 길이나 순서 및 구성에 따라 거리 값이 크게 달라질 수 있으므로, 결과를 해석할 때 이러한 특성을 함께 고려할 필요가 있다. 이러한 특성으로 인해 레벤슈타인 거리는 문자열 간의 차이를 정량적으로 비교할 필요가 있는 다양한 분야에서 활용될 수 있다. 예를 들어, 생물 정보학에서는 DNA나 단백질 서열 간의 유사성을 비교하는 데 사용되며, 음성 인식 과정에서도 음향 혼동성을 분석하는 기준으로 활용된다[8,9].

한편, 앞서 살펴본 거리 기반 방법들은 주로 각 차원에서의 값의 차이를 기반으로 두 객체 간의 유사성을 판단한다. 그러나 데이터의 크기보다는 방향이나 분포 형태의 유사성이 중요한 경우도 존재하며, 이러한 경우에는 거리 대신 벡터 간의 각도를 기반으로 유사성을 측정하는 방법이 사용될 수 있다.

2.4 코사인 유사도

코사인 유사도(Cosine similarity)는 두 벡터 간의 각도를 기반으로 유사성을 측정하는 방법으로, 벡터의 크기가 아닌 방향의 유사성에 초점을 맞춘다. n 차원 공간에 속한 두 벡터 $x = (x_1, x_2, \dots, x_n)$ 와 $y = (y_1, y_2, \dots, y_n)$ 에 대해 코사인 유사도는 다음과 같이 정의된다.

$$\cos(\theta) = \frac{(x \cdot y)}{\|x\| \|y\|} \quad (4)$$

식 (4)에서 $x \cdot y$ 는 두 벡터의 내적을 의미하며, $\|x\|$ 와 $\|y\|$ 는 각각 벡터의 크기(norm)를 나타낸다. 이 값은 두 벡터 사이의 각도 θ 의 코사인 값으로 해석되며, 일반적으로 값의 범위는 $[-1, 1]$ 사이에 존재한다.

코사인 유사도는 벡터의 크기 차이를 제거하고 방

향성만을 비교한다는 특징을 가지며, 동일한 방향을 가지는 경우 최대값을 갖고, 서로 직교하는 경우에는 0에 가까운 값을 갖는다. 이러한 특성으로 인해 각 요소의 절대적인 크기보다 상대적인 구성 비율이 중요한 경우에 유용하게 사용될 수 있다.

2.5 자카드 유사도

자카드 유사도(Jaccard similarity)는 두 집합 간의 유사성을 측정하는 방법으로, 두 집합의 교집합 크기를 합집합 크기로 나눈 비율로 정의된다. 집합 A와 B에 대해 자카드 유사도는 다음과 같이 표현할 수 있다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

식 (5)에서 $|A \cap B|$ 는 두 집합에 공통으로 포함된 원소의 개수를 의미하며, $|A \cup B|$ 는 두 집합에 포함된 전체 원소의 개수를 의미한다. 자카드 유사도의 값은 0과 1 사이의 값을 가지며, 두 집합이 완전히 동일할 경우 1, 공통 원소가 전혀 없을 경우 0의 값을 갖는다.

본 연구에서는 자카드 유사도를 텍스트 데이터에 적용하기 위해, 문자열을 해당 문자열을 구성하는 문자(character)들의 집합으로 표현하여, 두 문자열 간의 유사성을 계산한다. 즉, 문자열 s와 t에 대해 각각의 문자 집합을 $C(s)$, $C(t)$ 로 정의하고, 이를 기반으로 자카드 유사도를 다음 식 (6)과 같이 계산한다.

$$J(s, t) = \frac{|C(s) \cap C(t)|}{|C(s) \cup C(t)|} \quad (6)$$

이때 두 문자열 간의 유사도는 공통적으로 포함된 문자 집합의 크기를 전체 문자 집합의 크기로 정규화한 비율로 해석할 수 있다.

자카드 유사도는 원소의 존재 여부를 중심으로 정의되기 때문에, 데이터가 희소한 구조를 가지는 경우에도 안정적으로 적용될 수 있다. 이와 같은 특성으로 인해 자카드 유사도는 집합 기반 데이터 뿐 아니라 다양한 형태의 이산적 데이터 표현에서 두 대상 간의 유사성을 비교하는 데 효과적으로 활용될 수 있다. 텍스트에 적용하는 경우, 어순에 독립적이며 문맥적 변형에 강건하다는 장점이 있다.

3. 제안 방법론

체계 장비에 대한 데이터는 장비명, 길이, 중량, 기능 설명 등 서로 다른 형태의 데이터 항목으로 구성된다. 각 데이터 항목은 서로 자료형도 다르고 데이터의 분포도 다르므로, 단일 유사도 기법만으로 장비 간의 비교를 하는 데에 한계가 있다. 따라서 본 연구에서는 데이터 항목별 특성에 적합한 유사도 기법을 적용하고, 이를 통합하여 장비 간의 종합 유사도를 산출하는 방법을 제안한다. 이러한 종합 유사도를 통해 유사 장비를 식별하고 나아가 제안 단계에서 데이터를 기반으로 장비의 신뢰도를 예측할 수 있다.

3.1 무기체계 교범의 장비별 데이터 항목

3.1.1 장비명

장비명은 장비를 식별하기 위한 대표 명칭으로, 비교적 짧은 길이의 텍스트로 구성된다는 특징이 있다. 일반적으로 장비명에는 장비의 기능, 계열, 모델명, 운용 목적 등의 정보가 포함되며, 동일 계열 장비 간에는 공통적으로 사용되는 단어 또는 표현이 반복되는 경향이 있다.

3.1.2 물리적 제원

장비의 가로, 세로, 높이와 같은 길이 제원은 장비의 외형적 구조를 가장 직접적으로 나타내는 물리적 제원으로 별도의 시험 없이 쉽게 측정이 가능하다는 장점이 있다.

장비의 무게를 나타내는 중량 제원은 장비에 작용하는 중력의 크기를 나타내는 물리적 제원으로, 비교적 간단한 계측만으로 확보할 수 있다는 장점이 있다. 장비 중량은 장비를 구성하는 부품의 수, 재질, 내부 구성의 집적도 등에 따라 달라질 수 있으며, 장비 간 물리적 특성을 비교하는 데 유용한 기준이 될 수 있다.

3.1.3 장비기능

장비의 기능 정보는 해당 장비가 수행하는 역할, 운용 목적 및 주요 특성을 설명하는 텍스트로 구성된

다. 기능 정보는 장비명에 비해 상대적으로 긴 길이의 서술형 텍스트 형태를 가지며, 장비의 기술적 특징을 구체적으로 포함한다는 특징이 있다.

3.2 데이터 항목별 적정 유사도

2.1.(유클리디안 거리)~2.5.(자카드 유사도)에서 제시한 5가지 유사도 기법들에 대하여 데이터 항목별로 적절한 유사도 기법을 선택해야 한다. Table 1을 보면 알 수 있듯이 각 유사도 기법마다 적용 가능한 자료형이 있으며, 따라서 유사도 기법을 적용하기에 앞서서 적용할 데이터 항목의 자료형을 고려해야 한다. 본 연구의 데이터 항목별 적정 유사도 선정 과정에서는 실제 데이터가 아니라 가상의 데이터만을 사용한다.

Table 1. Comparison between similarity measures

Category	Basis	Data Type
Euclidean Distance	Overall numerical difference	Vector
Manhattan Distance	Sum of absolute differences across dimensions	Vector
Levenshtein Distance	Edit operations	String
Cosine Similarity	Vector orientation	Vector
Jaccard Similarity	Common elements	Set

3.2.1. 장비명 및 장비기능에 대한 적정 유사도

장비명과 장비기능은 텍스트 데이터에 해당한다. 텍스트는 문자열이면서 동시에 문자의 집합으로 간주할 수 있다. 따라서 문자열 또는 집합 자료형에 적절한 레벤슈타인 거리 또는 자카드 유사도를 적용한다. 자카드 유사도의 경우 유사도 계산시 데이터 자체의 길이가 반영되어 계산이 되나, 레벤슈타인 거리의 경우 데이터 자체의 길이가 반영되지 않는다. 따라서 레벤슈타인 거리와 자카드 유사도를 동등하게 비교하기 위해 레벤슈타인 거리를 적용할 때 문자열 길이로 거리값을 정규화한다.

텍스트 데이터는 같은 의미를 가지고 있어도 서로 다른 문장으로 표현될 수 있다. 본 연구에서는 Table 2와 같이 서로 의미적으로 동일한 두 문장을 구성하고, 이 두 문장 각각에 동일한 유사도 기법을 적용하였을 때, 두 유사도/거리 값의 차이를 살펴본다. 또한,

Table 3와 같이 서로 의미적으로 상이한 두 문장 각각에 대하여 동일한 유사도 기법을 적용하였을 때, 두 유사도/거리 값의 차이를 구한다.

Table 2. Examples of semantically equivalent sentences with different syntactic structures

No.	Trans. Type	Sentence A	Sentence B
1	Active/Passive	장비는 입력 신호를 처리한다.	입력 신호는 장비에 의해 처리된다.
2	Predicate Structural Variation	장비는 외부 환경 변화에 따라 자동으로 보정된다.	외부 환경 변화에 따라 장비가 자동 보정된다.

Table 3. Example of semantically non-equivalent sentences with different syntactic structures

Sentence A	Sentence B
장비는 입력 신호를 처리한다.	장비는 네트워크 연결 상태를 감시한다.

Table 4에서의 f_A , f_B 는 각각 A 문장과 B 문장에 대해 유사도 기법, f 를 적용하여 구한 유사도/거리 값을 나타내며, eq/non-eq는 A 문장과 B 문장이 서로 의미적으로 동일한지 여부를 나타낸다(eq: 동일, non-eq: 동일하지 않음). Table 4에서 Δ 는 동일 문장과 비동일 문장 간의 유사도/거리값의 차이(%)를 나타낸다. Table 4를 보면 알 수 있듯이, 자카드 유사도는 레벤슈타인 거리에 비하여 Δ 가 더 극명하게 나타난다. 이에 따라 본 연구에서는 장비명 및 장비기능 데이터 항목에 자카드 유사도를 적용한다.

Table 4. Comparative results of jaccard similarity and levenshtein distance for equivalent and non-equivalent sentences

f	$ f_A - f_B _{eq}$	$ f_A - f_B _{non-eq}$	Δ (%)
Levenshtein Distance	0.33 (↓)	0.43 (↑)	30.3
Jaccard Similarity	0.34 (↑)	0.13 (↓)	61.8

3.2.2 길이 및 중량에 대한 적정 유사도

길이와 중량은 실숫값을 가지는 수치형 데이터에 해당하므로 유클리디안 거리, 맨해튼 거리, 또는 코사인 유사도를 적용한다.

길이 및 중량 정보를 기반으로 유사 장비를 판단할 때, 가로:세로:높이:중량의 비율이 아니라 실제 물리적 규모의 차이를 고려하는 것이 중요하다. 따라서 코사인 유사도보다는 맨해튼 거리나 유클리디안 거리가 더 적절하다.

Table 5. Recommended similarity and distance metrics according to data attribute characteristics

Category	Similarity/Distance
equipment name	Jaccard Similarity
equipment function	Jaccard Similarity
equipment size	Euclidean Distance
equipment weight	Manhattan Distance

3.3 데이터 항목 통합 유사도 산출

본 연구에서는 장비의 물리적 유사성을 비교하기 위하여 가로, 세로, 높이 및 중량 정보를 활용한다. 가로, 세로 및 높이는 장비의 공간적 형상을 구성하는 요소이며, 장비의 배치 방향 또는 회전에 따라 축 방향 정의가 달라질 수 있으므로 개별 차원값의 차이보다는 전체적인 공간적 규모 차이를 평가하는 것이 적절하다. 이에 따라 공간 형상 유사성을 비교하기 위해 식 (8)과 같이 가로, 세로 및 높이 값을 결합하여 차원 수 3인 벡터를 만들고 이에 유클리드 거리를 적용하여 형상에 대한 유사도를 계산한다. 반면, 중량값은 장비의 구조적 복잡도, 내부 구성 및 재질 특성을 반영하는 독립적인 물리량이므로 별도의 유사도 항목으로 분리한다. 유클리디안 거리와 맨해튼 거리 모두 유사도 계산에 적용할 수 있도록 역수 함수로 변환하여 수식을 구성한다. i 번째 장비와 j 번째 장비 간의 유사도 $sim_{i,j}$ 는 다음 식 (7)과 같이 정의한다.

$$\begin{aligned}
 sim_{i,j} = & \lambda_1 * jaccard(n_i, n_j) \\
 & + \lambda_2 * jaccard(f_i, f_j) \\
 & + \lambda_3 * \frac{1}{1 + euclidean(\mathbf{X}_i, \mathbf{X}_j)} \\
 & + \lambda_4 * \frac{1}{1 + manhattan(w_i, w_j)}
 \end{aligned} \quad (7)$$

$$\mathbf{X}_i = [x_i; y_i; z_i] \quad (8)$$

i 번째 장비에 대하여 n_i 는 장비명, f_i 는 장비기능, x_i , y_i , z_i 는 각각 가로, 세로, 높이, w_i 는 중량을 나타낸다. $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 는 각 유사도 항목에 대한 가중치 하이퍼파라미터를 나타낸다.

4. 적용 예제

본 장에서는 앞서 제안한 유사도 산출 방법론을 실제 사업 데이터를 참고하여 구축한 데이터베이스(DB)에 적용한 예를 제시한다. 적용 예제의 목적은 신규 장비에 대해 과거 장비 데이터와의 유사도를 정량적으로 산출하고, 이를 기반으로 유사 장비를 우선적으로 식별할 수 있음을 보이는 데 있다. 이를 위해 과거 수행 사업 중 2개 사업에서 확보한 장비 데이터를 참고하여 DB를 구축하고, 가상의 분석 대상 장비를 1종에 대해 Top 5 유사장비를 도출하는 절차를 수행하였다.

4.1 분석 데이터베이스 구축

적용 예제를 위한 분석 데이터베이스는 과거 수행 사업에서 활용된 장비 데이터의 구조와 항목 체계를 참고하여, 보안성을 고려한 임의 데이터셋 형태로 구축하였다. 실제 무기체계 사업 자료에는 장비 구성, 제원, 기능 등 민감할 수 있는 정보가 포함될 수 있으므로, 본 연구에서는 실데이터를 직접 제시하지 않고 장비 데이터의 형식과 분포 특성만을 반영한 가상 데이터를 생성하여 활용하였다.

DB에 포함된 데이터 항목은 장비명, 제원(가로, 세로, 높이, 중량), 기능으로 구성하였다. 이는 실제 교범 및 장비 목록 자료에서 공통적으로 식별 가능한 기본 속성을 기준으로 선정한 것이며, 유사도 분석에 필요한 최소 단위 정보만 반영하였다. 다만 각 장비명 및 기능 설명은 실제 장비를 직접 식별할 수 없도록 일반화하였고, 제원 정보 또한 실측값을 그대로 사용하는 대신 분석 목적에 적합하도록 범위와 분포를 고려하여 조정하였다.

또한 본 연구에서 사용한 장비 데이터는 특정 체계나 사업을 재현하기 위한 것이 아니라, 제안한 유사도 산출 방법론의 적용 가능성을 설명하기 위한 예시 데이터로 구성되었다. 이를 위해 전원, 제어, 변환, 보호, 분배 등 유사성이 일부 중첩될 수 있는 장비군을

포함하도록 설계하였으며, 장비 간 유사도 차이가 단계적으로 나타날 수 있도록 장비명, 제원, 기능 요소를 조합하였다. 특히 분석 대상 장비와 높은 유사도를 갖는 일부 장비와, 동일 계열이지만 세부 기능이나 형상이 다른 장비를 함께 포함시켜 top-N 유사장비 도출 결과가 명확히 구분되도록 하였다.

최종적으로 구축된 DB는 각 장비를 하나의 레코드로 관리하는 구조로 구성하였으며, 신규 장비 입력 시 기존 장비군 전체와의 유사도를 일괄적으로 산출할 수 있도록 설계하였다. 이러한 방식은 보안 제약이 있는 환경에서도 제안 방법론의 절차와 활용 가능성을 검토할 수 있다는 점에서 실용적 의미가 있다.

4.2 분석 대상 장비 설정

유사도 산출 예시를 위해 가상의 신규 장비 1종을 분석 대상으로 설정하였다. 해당 장비는 “전원분배장치” 계열의 장비로 가정하였으며, 주요 속성은 다음과 같다.

장비명: 전원분배장치

제원: 가로 520 mm, 세로 430 mm, 높이 210 mm, 중량 28 kg

기능: 외부 전원을 입력받아 다수의 하위 구성품에 전원을 분배하고, 과전류 보호 및 상태 표시 기능 수행

해당 장비는 기존 DB 내 여러 전원공급계통 장비와 일부 특성이 유사할 수 있으나, 명칭이 완전히 동일하지 않거나 세부 기능이 부분적으로만 일치할 가능성이 존재한다. 따라서 단일 항목이 아니라 복수 항목을 종합적으로 반영하는 가중합 기반 유사도 산출 방식의 유효성을 확인하기에 적절한 사례로 볼 수 있다.

4.3 유사도 산출 절차

적용 절차는 다음과 같다. 먼저 신규 장비의 장비명, 제원, 기능 데이터를 입력한다. 이후 DB에 저장된 각 과거 장비와 항목별 유사도를 산출한다. 장비명은 문자열 유사도 기법을 적용하여 명칭의 형태적 유사성을 평가하였고, 제원 정보는 수치형 거리 기반 기법을 적용하여 크기 및 중량의 근접성을 계산하였다. 기

능은 서술형 텍스트의 의미적 유사성을 반영할 수 있는 방식으로 유사도를 산출하였다.

이후 각 항목별 유사도에 사전에 정의한 가중치를 반영하여 최종 유사도 점수를 계산하였다. 예를 들어 기능 항목은 장비의 실질적 역할을 나타내므로 상대적으로 높은 가중치를 부여하고, 제원은 물리적 호환성 및 설치성 측면을 반영하므로 그 다음 수준의 가중치를 적용할 수 있다. 장비명은 명칭 관례의 영향을 받을 수 있으므로 비교적 낮은 가중치를 두되, 동일 계열 식별에는 유효한 보조 지표로 활용할 수 있다.

가중치 하이퍼파라미터는 분석대상 체계의 특성을 고려하여 사용자가 적절한 값을 설정함이 권장되나, 본 예제에서는 설명의 편의를 위해 장비명 0.2, 제원 0.3, 기능 0.5의 가중치를 적용하였다. 최종 유사도 점수는 식 (9)와 같이 산출된다.

$$S_{final} = 0.2 \times S_{name} + 0.3 \times S_{spec} + 0.5 \times S_{function} \quad (9)$$

여기서 S_{name} 은 장비명 유사도, S_{spec} 은 제원 유사도, $S_{function}$ 은 기능 유사도를 의미한다.

4.4 TOP 5 유사장비 도출 결과

가상의 신규 장비에 대해 DB 전체를 대상으로 유사도를 산출한 결과, 상위 5개 장비는 다음과 같이 도출되었다.

Table 6. Top 5 similar equipment identified for the target equipment

rank	equipment name	name sim	spec sim	func sim	final sim
1	전원공급분배장치	0.75	0.878	0.889	0.858
2	전원분배기	0.571	0.923	0.861	0.822
3	전원보호분배기	0.444	0.799	0.838	0.747
4	전원공급기	0.222	0.725	0.757	0.640
5	전원제어함	0.222	0.799	0.625	0.597

유사도 산출 결과, 분석 대상 장비와 가장 유사한 장비는 전원공급분배장치(0.858)로 나타났다. 해당 장비는 장비명, 제원, 기능 전 항목에서 고르게 높은

유사도를 보여, 분석 대상과 가장 근접한 장비로 판단되었다. 특히 제원 유사도(0.878)와 기능 유사도(0.889)가 높게 산출되어 물리적 특성과 운용 목적 측면에서 높은 일치성을 확인할 수 있었다.

다음으로 전원분배기(0.822)가 2순위로 도출되었다. 이 장비는 제원 유사도(0.923)가 가장 높게 나타나 분석 대상 장비의 제원과 매우 유사한 것으로 분석되었다. 다만 장비명 유사도는 상대적으로 낮아 최종적으로는 1순위보다 낮은 점수를 보였다. 이는 본 방법론이 명칭뿐 아니라 기능적 유사성을 중요하게 반영함을 보여준다.

3순위인 전원보호분배기(0.747)는 제원 및 기능 측면에서 비교적 높은 유사도를 보였으나, 장비명 유사도가 낮아 상위 2개 장비보다는 낮은 순위를 기록하였다. 4순위 전원공급기(0.640)와 5순위 전원제어함(0.597)은 전원 계열 장비라는 공통성은 있으나, 각각 공급 중심 및 제어 중심의 성격이 상대적으로 강해 최종 유사도가 낮게 산출되었다.

4.5 결과 해석 및 활용 가능성

본 적용 예제를 통해 제안 방법론은 단순 명칭 일치 여부를 넘어서, 물리적 제원과 기능 정보를 함께 반영함으로써 신규 장비와 유사한 과거 장비를 보다 합리적으로 식별할 수 있음을 확인하였다. 특히 명칭이 일부 상이한 경우에도 기능 및 제원 기반 유사도가 이를 보완하여, 실제 검토 가치가 높은 후보 장비를 상위에 도출할 수 있다는 점에서 실무적 활용성이 있다.

이러한 결과는 신규 장비 도입 시 기존 장비와의 유사성을 기반으로 정비개념, 군수지원요소, 시험평가 항목, 교범 재활용 가능성 등을 사전에 검토하는 데 활용될 수 있다. 또한 사업 초기 단계에서 유사 장비 군을 신속히 식별함으로써, 과거 사업의 산출물과 연계한 분석 효율 향상에도 기여할 수 있다.

5. 결론

본 연구에서는 무기체계 신뢰도 분석의 유사장비 비교법에 적용할 수 있는 정량적 유사도 산출 방법을 제안하였다. 기존 유사장비비교법은 과거 개발 및 운용 장비의 신뢰도 데이터를 활용할 수 있다는 장점이 있으나, 유사성 판단 과정이 분석자의 경험과 직관에

의존한다는 한계가 존재하였다. 또한 사업 수행 사례가 축적됨에 따라 방대한 데이터 중 적절한 유사 장비를 탐색하는 과정 역시 점차 비효율적으로 변하고 있다. 이에 본 연구에서는 정보공학 및 데이터 분석 분야에서 활용되는 다양한 유사도 측정 기법을 무기체계 데이터에 적용하고, 장비 특성별로 적합한 기법을 조합하여 통합 유사도를 산출하는 방법론을 제안하였다.

이를 위해 유클리디안 거리, 맨해튼 거리, 레벤슈타인 거리, 코사인 유사도, 자카드 유사도 등 대표적인 유사도 및 거리 기반 기법의 특성을 분석하고, 장비명, 제원, 기능과 같은 데이터 항목의 특성에 따라 적절한 기법을 적용할 수 있음을 확인하였다. 또한 항목별 중요도를 반영하기 위하여 가중합 기반의 통합 유사도 산출 방식을 적용하였으며, 가상의 데이터를 기반으로 구축한 DB와 가상의 신규 장비 사례를 통해 적용 가능성을 검토하였다.

적용 결과, 제안 방법론은 단순한 장비명 일치 여부를 넘어서 기능 및 물리적 제원을 함께 고려함으로써, 실제 운용 및 정비 관점에서 유사성이 높은 장비를 효과적으로 식별할 수 있음을 확인하였다. 특히 명칭이 상이하더라도 기능과 제원 기반 유사도가 높은 장비를 상위 후보로 도출할 수 있었으며, 이는 기존 방식 대비 보다 합리적이고 재현성 있는 유사장비 선정이 가능함을 보여준다. 또한 다수의 과거 사업 데이터를 대상으로 유사도 산출 과정을 자동화할 수 있다는 점에서, 분석 효율성과 활용성 측면에서도 의미가 있다.

본 연구에서 제안한 방법론은 향후 신규 무기체계 개발 시 유사 장비 기반의 신뢰도 예측, 정비개념 검토, 군수지원요소 분석, 시험평가 항목 도출 등 다양한 RAM 업무에 활용될 수 있다. 또한 사업 초기 단계에서 과거 개발 사례와의 연계를 지원함으로써, 제한된 정보 환경에서도 보다 객관적인 의사결정을 수행하는 데 기여할 수 있을 것으로 기대된다.

나아가 향후 실제 적용 사례를 확대하고 인공지능 자연어처리를 통하여 의미 기반의 고도화된 유사도 분석 체계로 확장될 수 있다.

참고문헌

- [1] 방위사업청, 무기체계 RAM업무지침, 2021.

- [2] Byung-ok Kil, 'Study on the Establishment Plan of RAM Analysis Model for the Defense Weapon System,' Journal of Northeast Asian Studies, VOL. 23, NO. 1, 2018, pp. 77–98.
- [3] DOD, MIL-HDBK-217F Notice 2, 1991.
- [4] DOD, MIL-HDBK-338B, 1998.
- [5] Seung-Jin Han & Soonhuem Paik, 'RAM Goal-Setting Using the AHP and Field Data of Similar Weapon Systems,' Journal of Applied Reliability, VOL. 20, NO. 2, 2020, pp. 154–162.
- [6] 서영진·박경준, 'BOM 데이터베이스를 활용한 무기체계 유사 장비 분석모델,' 2024년도 한국통신학회 하계종합학술발표회 논문집, 2024.06, pp. 1751–1752.
- [7] Mario Köppen, 'The Curse of Dimensionality,' in proceedings of the 5th Online World Conference on Soft Computing in Industrial Applications (WSC5), VOL. 1, 2000.
- [8] Brian Young, Tom Faris, & Luigi Armogida, 'Levenshtein Distance as a Measure of Accuracy and Precision in Forensic PCR-MPS Methods,' Forensic Science International: Genetics, VOL. 55, 2021, Article 102594.
- [9] A. Zgank & Z. Kacic, 'Predicting the Acoustic Confusability Between Words for a Speech Recognition System Using Levenshtein Distance,' Elektronika Ir Elektrotechnika, VOL. 18, NO. 8, 2012, pp. 81–84.