Journal of The Korea Society of Computer and Information
Vol. 24 No. 7, pp. 125-133, July 2019
www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.07.125

Finding Naval Ship Maintenance Expertise Through Text Mining and SNA

Jin-Gwang Kim*, Soung-woong Yoon*, Sang-Hoon Lee*

Abstract

Because military weapons systems for special purposes are small and complex, they are not easy to maintain. Therefore, it is very important to maintain combat strength through quick maintenance in the event of a breakdown. In particular, naval ships are complex weapon systems equipped with various equipment, so other equipment must be considered for maintenance in the event of equipment failure, so that skilled maintenance personnel have a great influence on rapid maintenance. Therefore, in this paper, we analyzed maintenance data of defense equipment maintenance information system through text mining and social network analysis(SNA), and tried to identify the naval ship maintenance expertise. The defense equipment maintenance information system is a system that manages military equipment efficiently. In this study, the data(2,538cases) of some naval ship maintenance teams were analyzed. In detail, we examined the contents of main maintenance and maintenance personnel through text mining(word cloud, word network). Next, social network analysis(collaboration analysis, centrality analysis) was used to confirm the collaboration relationship between maintenance personnel and maintenance expertise. Finally, we compare the results of text mining and social network analysis(SNA) to find out appropriate methods for finding and finding naval ship maintenance expertise.

▸Keyword: Collaboration relationship, Expert, Social Network Analysis(SNA), Text Mining

I. Introduction

1.1 The Purpose of Study

학술연구 분야에서 저자와 초록, 키워드 등을 분석하여 특정 분야의 지식구조를 파악하거나 전문가 검색 및 추천 등에 관한 많은 연구가 진행되어왔다. 그리고 최근에는 소셜 네트워크 서비스(이하 SNS) 등장과 더불어 학술연구 분야 외에도 SNS상의 '질의-답변' 관계에서 답변의 평점을 이용하여 전문가를 검색하는 방법도 연구되고 있다[1]. 하지만 일반적으로 학술연구는 다소 시간이 소요되기에 저자들의 연구 활동은 수 건에서 수십 건 정도이며, 그 결과 분석에 쓰이는 데이터가 많지 않을 경우에는 의미 있는 패턴 또는 결과 확인이 제한될 수도 있다. 그리고 학술연구 활동이나 SNS를 하지 않는 현장 작업자 또는 실무자의 경우에도 전문가 식별이 역시 제한된다.

그런데도 이처럼 전문가 검색 등에 관한 연구가 지속적으로 필요한 이유는 전문가들이 자신의 분야에서 다년간의 유사경험을 통한 통찰력을 기반으로 신속한 업무처리 및 조언 등을 통해 조직에 기여하는 역할이 크기 때문이다. 따라서 조직 내 분야별 전문가를 확인하는 것은 예상치 못한 문제에 봉착했을 때 신속한 문제 해결에 큰 도움이 되기에 효율적인 조직 운영을 위해서는 필수적이다. 그러나 국방 분야에서는 보안을 고려하는 군사자료의 특성상 이를 분석하여 전문가를 검색 및 추천하는 등에 관한 연구가 거의 이루어지지 못했다. 또한, 주기적인 전속으로 담당 업무가 자주 변경되어 분석을 위한 자료들이 없거나 찾기 힘든 실정이다. 그래도 일부 국방망(군 내부망)을 이

• First Author: Jin-Gwang Kim, Corresponding Author: Sang-Hoon Lee
*Jin-Gwang Kim (whiteseas@naver.com), Dept. of Computer Science, Korea National Defense University
*Soungw-Woong Yoon (ysw1209@gmail.com), Dept. of Computer Science, Korea National Defense University
*Sang-Hoon Lee (hoony@kndu.ac.kr), Professor. Dept. of Computer Science, Korea National Defense University
• Received: 2019. 04. 30, Revised: 2019. 06. 24, Accepted: 2019. 06. 26.

용한 시스템 등에는 데이터가 수년간 축적되어 있어 이를 기반으로 전문가 탐색이 가능하며, 아래에서는 다양한 국방 분야 중 현대전에서 전쟁의 승패에 큰 영향을 미치는 무기체계의 성능 보장과 관련된 군 정비 전문가에 관해 확인하려 한다.

무기체계가 아무리 성능이나 내구성이 좋더라도 고장 후 수리가 되지 않는다면 그 무기체계는 더 이상 사용가치가 없으며, 따라서 고장 시 신속한 장비복구를 통한 전투력 보존은 매우 중요하다. 특히, 규모가 큰 무기체계인 해군 함정은 사격통제, 탐지체계 등 다양한 장비들이 탑재 운용되는 복합 무기체계로 장비 손상 시 복구를 위해 관련된 다른 장비들까지도 고려해야 하므로 경험 있는 전문 정비자가 신속한 장비복구에 미치는 영향이 매우 크다. 그리고 일부 정비반별로 조금 차이는 있겠지만 개인적인 경험에 비추어 볼 때 복잡하고 어려운 정비의 경우에는 해당 장비 전문 정비자를 중심으로 정비가 진행됨을 수차례 경험했다. 즉, 정비경력이 많더라도 자신의 주 정비 분야가 아니거나 해당 장비의 정비 경험이 없다면 장비의 전원계통, 작동원리 등에 대한 세부적인 이해도가 낮아 수리가 제한되었으며, 이는 규모가 큰 함정을 신속하게 정비하기 위해 정비자별로 분업 정비하는 정비부대의 특성에서 기인했다고 할 수 있다. 따라서 본 연구에서는 국방 장비정비정보체계 내 축적된 정비데이터 기반으로 정비반별 또는 분야별 정비 전문가를 탐색하여 필요 시 신속한 장비복구와 원격 정비상담자 등으로 해당 전문인력을 활용하고자 한다.

1.2 The Scope and Method of Study

특수한 목적을 위해 사용하는 군 무기체계들은 일반 상용제품과 비교하면 소규모 다품종으로 국내·외에서 주문 제작되기에 A/S가 절대 쉽지 않다. 비용도 많이 들지만, 때에 따라서는 장비 제작업체가 도산하여 수리가 제한되는 경우도 비일비재하다. 따라서 군은 장비를 효율적으로 정비 및 관리하기 위해 정비관리시스템(가칭) 등을 운용해왔으며, 2009년부터는 각 군별로 운용 중이던 정비관리시스템을 통합하여 각종 무기체계, 수리 부속 및 공구의 보급 등 정비 관련사항을 관리하는 국방 장비정비정보체계를 운용 중이다. 국방 장비정비정보체계에는 모든 정비의 시작부터 완료까지 연간 수백만여 건의 정비데이터(정비내용, 정비시간 등)가 기록되고 있으며, 본 연구에서는 2017년 해군 정비데이터 중 특정 정비반에 해당하는 일부(2,538건)를 데이터마이닝 기법을 적용하여 군 정비 전문가를 파악하고자 한다. 세부적인 연구방법으로는 텍스트마이닝을 통해 정비반의 주요 정비내용과 정비자 현황을 살펴보고, 사회연결망 분석을 통해 정비자 간 협업 관계, 정비자의 전문분야 등을 확인하여 이를 종합 분석함으로써 전문가 탐색 및 탐색에 적합한 방법을 제시하고자 한다.

II. Related works

2.1 Expert & Knowledge Management System

전문가 탐색에 앞서 전문가들이 가진 전문지식을 활용하기 위한 연구 및 노력은 수십 년 전부터 계속되어왔으며, 대표적으로 전문가시스템과 지식경영시스템이 있다. 그리고 전문가 탐색에 관한 연구는 전문가의 지식과 경험 등을 활용하는 측면에서 전문가 및 지식경영 시스템과 유사한 연구 분야로 볼 수 있다. 각 시스템에 대해서 간략하게 살펴보면 전문가시스템(Expert System)이란 전문가의 전문지식을 수집, 정리하여 주어진 특정 전문영역에 관한 문제를 해결하는 시스템이다. 활용 분야로는 의료 진단, 고장 진단, 경영 등 인간의 지적 능력이 필요로 하는 분야에 적용되고 있다[2]. 다음으로 지식경영 시스템(Knowledge Management System)은 지식관리 시스템이라고도 하며 조직지식을 관리하기 위해 응용되는 정보시스템의 유형을 의미한다. 조직 내에 축적되는 각종 지식과 Know-how를 효율적으로 관리하고 이를 상호 공유할 수 있도록 하며 데이터베이스 검색을 통해서 전문가를 찾거나 혹은 지식의 기록된 원천을 찾는 등에 활용되기도 한다[3]. 목적과 활용 분야에서 두 시스템을 비교하면 <표 1>과 같으며, 의미는 조금 다르나 결국 지식과 경험 등을 축적 관리하여 활용한다는 점에서 상호 유사한 시스템이라 할 수 있겠다. 다만 전문가 탐색은 시스템을 구축하여 활용하는 것이 아닌 역으로 시스템의 기반이 되는 지식과 경험의 출처, 제공자를 확인한다는 부분에서 조금 차이가 있다[4].

Table 1. Expert & Knowledge Management System

	Expert System	Knowledge Management System
Purpose	Collecting and organizing knowledge to solve problems	Manage and mutually share knowledge and know-how
Field of application	Medical and trouble diagnosis, management plan, etc.	Work sharing and collaboration, knowledge reuse, etc.

2.2 Data Mining(Text mining & SNA)

데이터마이닝은 통계학, 기계학습, 컴퓨터과학 등의 학문이 연관된 분야로 다양하게 정의될 수 있으며, 일반적으로 대규모 데이터에서 가치 있는 정보를 추출하는 것을 말한다. 즉, 의미 있는 패턴과 규칙을 발견하기 위해서 자동화되거나 반자동화된 도구를 이용하여 대량의 데이터를 탐색하고 분석하는 과정이다[5]. 데이터마이닝은 빅데이터 기술 발전과 더불어 발전하고 있으며, 분석방법에는 텍스트마이닝, 오피니언마이닝, 통계적 분석, 소셜 네트워크 분석 등이 있다. 본 연구에서는 텍스트마이닝과 소셜 네트워크 분석을 사용할 것이며 간략하게 살펴보면 아래와 같다.

텍스트마이닝이란, 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치 있고 의미 있는 정보를 찾아내는 기법으로 정보 간의 연계성을 파악하고 텍스트 내 잠재

된 토픽을 찾아내는 등 다양한 결과를 도출할 수 있다[6].

사회연결망 분석은 노드들의 상호 연결 관계를 네트워크 분석을 통해 알아내는 방법으로 복잡하게 얽혀있는 관계를 한 눈에 시각화하여 조금 더 쉽게 설명할 수 있는 도구이다. 최근 사회연결망 분석의 범위는 인간관계를 넘어 단어 사이의 관계로까지 확장되고 있으며 컴퓨터 계산 속도와 시각화 방법의 발전으로 이전까지 고려하기 어려웠던 관계와 그들 간 상호작용을 이해하고 분석할 수 있게 되었다. 또한, 사회연결망 분석의 연결망 그래프는 분석적 의미를 표현할 뿐만 아니라, 시각화된 패턴을 통해 더 많은 해석의 가능성을 열어준다. 연결망 그래프를 잘 표현하면 복잡하게 얽혀있는 연결망 구조를 한눈에 파악하여 중심과 고립 부분 등을 쉽게 확인할 수 있다[7].

2.3 General Information of Navy Ship Maintenance

해군 함정은 무기체계임과 동시에 부대이며 장병들이 근무하고 생활하는 공간이다. 또한, 함정 1척이 곧 전투력을 의미하기 때문에 효율적이고 체계적인 관리를 위해 연 2회 정기적으로 정비를 실시한다. 간략하게 해군 함정정비에 대해서 알아보면, 함정정비란 함정의 선체와 장비의 성능유지 및 수명연장을 위해서 손질, 검사, 수리, 재생, 개조, 교정하는 등의 일체 행위를 말한다.

정비의 종류에는 부대정비, 야전정비, 창정비 및 상가정비로 총 4가지가 있으며, 부대정비는 함정 승조원에 의해서 수시로 실시하는 정비이고, 야전정비는 함정 승조원들의 능력을 초과하는 수리사항을 정비지원 부대에서 실시하는 정비이다. 창정비는 함정 전반에 걸친 검사, 정비 수리사항을 정비지원 부대에 의해 실시하는 정비로 필수교환 품목 등을 교환한다. 마지막으로 상가정비는 수면 하 선체 및 구조물의 검사수리를 위해 상가시설을 구비한 정비지원 부대 또는 민간업체에 의해 실시하는 정비이다[8]. 이처럼 정비 규모 및 수준 등에 따라 정비종류가 나뉘며, 일반적인 정비 진행 과정은 아래 <그림 1>과 같다.

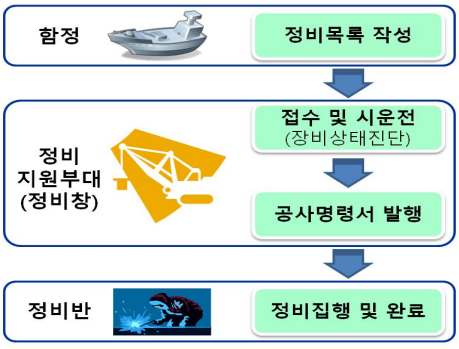

Fig. 1. Navy Ship's Maintenance Process

III. The Proposed Scheme

본 연구에서는 해군 장비정비정보체계의 다양한 정비데이터 중 정비관리번호, 정비내용 등 20개 속성을 포함하는 '공사명령서 발행현황' 릴레이션(이하 공사명령서)과 정비자, 정비관리번호 등 18개 속성을 포함하는 '일일 소비공시' 릴레이션(이하 일일 소비공시)을 대상 데이터로 선정하였다. 특히, 일일 소비공시는 장비정비정보체계의 일반 사용자 메뉴에는 나타나지 않는 데이터로 체계 관리자도 조회만 하는 수준으로 활용실적이 낮았다. 하지만 정비 관련 전문가를 찾기 위해서 '어떤 정비자가 어떤 장비의 정비에 참여했는지'에 대한 세부적인 확인이 필요하므로 일일 소비공시 데이터가 필요하였고, 공사명령서와 정비관리번호 속성으로 외부조인을 실시 후 데이터마이닝(텍스트마이닝, SNA)을 진행하였다.

3.1 Data Collection and Preprocessing

정비 전문가를 찾는데 필요한 데이터에 대해 생각해보면 기본적으로 정비자, 정비내용 등의 속성값을 포함해야 한다. 그러나 국방 장비정비정보체계는 전문가 탐색을 위한 시스템이 아니기에 별도로 해당 데이터 릴레이션은 존재하지 않았다. 따라서 데이터베이스 내 존재하는 릴레이션들을 통해서 연구에 필요한 릴레이션을 생성해야 했으며, 이는 정비한 내용을 포함하는 '공사명령서'와 정비자가 참여한 정비번호가 담겨있는 '일일 소비공시'를 외부조인함으로써(MS ACCESS 이용) 원하는 정비자, 정비내용 속성 값 등을 갖는 릴레이션(이하 세부 정비내역)을 얻을 수 있었다.

수집한 데이터는 해군 정비부대 중 규모가 가장 큰 ○○창의 최근 5년간(2013년~2017년) 정비데이터이며, 본 연구에서는 2017년 정비데이터를 분석 대상으로 선정하였다. 이후 분석 전 데이터 이상값 확인결과 2017년 ○○창 '세부 정비내역'의 정비데이터(총 426,627건) 중에 정비자의 정비내용을 확인할 수 없는 결측값 데이터가 있었다. 결측값은 장비정비정보체계 이용 시 입력자 또는 체계의 일시적인 오류, 점검 등에 의한 것으로 해당 데이터는 분석이 불가하기에 제외하고 약 44.6% 감소한 총 236,163건으로 정리하였다. 또한, 수십만 건의 데이터를 한 번에 분석하는 것은 제한되며 연구를 통해 확인하고자 하는 것은 정비반별 분석을 통한 정비 전문가 또는 정비자의 전문 정비 분야 확인이므로 정비반별로 데이터를 분류하였고, 77개(2017년 정비데이터 기준)의 정비반 중 임의로 '○○P반' 선정하여 2,538건의 정비데이터에 대한 분석을 진행하였다. 그리고 정비데이터는 일반자료로 분류되나 정비자 군번, 탑재 장비명 등의 개인정보와 군사정보가 담겨있어 불필요한 정보유출 방지와 보안을 위해 엑셀의 VBA(Visual Basic For Application) 이용 데이터 치환 작업을 하였다.(치환 예 : 연료유정유기 → 연*유*유*)

3.2 Visualization of Data Relationship

정비데이터 관계 시각화를 위해 '세부 정비내역'의 38개 속성 중 5개 값(정비자, 정비관리번호, 장비명, 정비내용, 정비시간)을 선택하여 Netminer(4.4.1.c) [9]를 이용 텍스트마이닝과

SNA를 실시하였다. 한 정비반에서 특정 정비자가 자주 등장한다는 것은 잦은 정비 빈도를 의미하며, 이는 정비 경험으로 볼 수 있다. 따라서 텍스트마이닝과 SNA를 실시하면 빈도 수가 높고 중심이 되는 정비자, 정비내용을 확인할 수 있으며, 상호 연관성을 파악한다면 정비 전문가 탐색이 가능할 것이다.

3.2.1 Centric-based Text Mining

텍스트마이닝 기법은 일반적으로 비정형 텍스트 데이터로부터 단어를 추출하고 문서, 문단, 문장 내의 단어 간의 인접성을 기준으로 네트워크를 구성할 수 있다. 이러한 텍스트 분석 기법을 이용하여 문서 간 연관성, 텍스트에 등장한 인물들의 관계를 시각화할 수도 있다. 본 연구의 정비데이터는 기본적으로 정형 데이터이므로 '정비자-정비내용' 속성 값을 병합하여 분석을 진행하였으며, 아래 수식의 TF-IDF 기반 워드 클라우드 및 네트워크로 시각화할 수 있다.

$$TF-IDF = TF \times IDF = TF \times \log\left(\frac{N}{DF}\right)$$

TF : 문서 내 특정단어 빈도, *N* : 전체 문서 수
DF : 여러 문서 내 특정단어 빈도, *IDF* : DF의 역수
< Term Frequency - Inverse Document Frequency >

3.2.2 Visualization using SNA

SNA는 노드 간의 방향성과 가중치를 연구자가 지정할 수 있으므로 빈도 수에 따라 자동으로 노드를 추출하는 텍스트마이닝과는 차이가 있다. 정비자 간 실시한 정비데이터 기반으로 협업 관계를 파악하고, 연결망 상에서 얼마나 중심적인 위치에 있는지를 나타내는 중심성 분석 중 위세중심성과 페이지랭크 2가지 방법으로 시각화하였으며, 각 중심성 값은 아래 수식에 따라 측정된다[10].

$$C_e(V_i) = \frac{1}{\lambda}\sum_{j=1}^{n} A_{j,i} C_e(V_i)$$

Ce(Vi) = 위세중심성 값, *λ* = 고정상수(고유벡터 값)
Aj,i = 인접행렬(n x n)
< EigenVector Centrality >

$$C_p(V_i) = \alpha\sum_{j=1}^{n} A_{j,i}\frac{C_p(V_i)}{d_j out} + \beta$$

Cp(Vi) = 페이지랭크 값, *α* = 감쇠계수(damping factor)
djout = out 링크의 수, *β* = 상수 값(또는 1-α/N(노드 수))
< PageRank Centrality >

위세중심성은 노드에 연결된 링크의 수뿐만 아니라 연결된 주변 노드의 영향력까지 반영해서 자신의 중심성을 계산하며, 페이지랭크는 위세중심성의 변형으로 연결된 주변 노드의 영향력을 참조하는 링크의 수로 나누어 영향력이 지나치게 퍼지는 것을 막는다. 비록 페이지랭크가 위세중심성의 일부 문제점을 보완한 것이긴 하지만 상황에 따라 쓰임이 다를 수 있으므로 각 연구 목적에 적절한 중심성 방법을 사용해야 한다. 본 연구에서는 2가지 중심성 결과를 비교 분석하여 전문가 탐색결과를 검증 및 적합한 중심성 방법을 확인하고자 한다.

IV. Analysis Results

4.1 Result of Text Mining

텍스트마이닝 적용을 통해 자동 추출된 노드들의 워드 클라우드는 <그림 3>와 같으며, 큰 글자의 노드가 빈도수가 높고 중요한 노드이다.

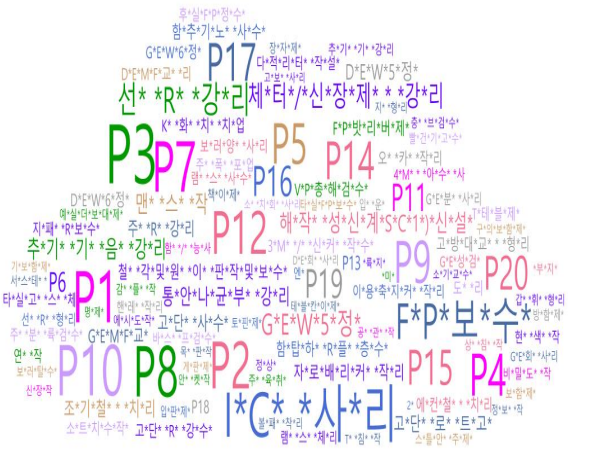

Fig. 3. Word Cloud(Maintenance detail)

<그림 3>을 살펴보면 P1, P2 등은 정비자를 의미하고, '*' 처리된 값은 치환된 정비내용을 나타낸다. 분석 대상인 '○○P반'에는 P1부터 P20까지 총 20명의 정비자가 있으며, 그 중 P3, P10의 정비자가 많은 정비에 등장했고, 정비내용에는 'I*C* *사*리'와 'F*P*보*수*'가 많음을 알 수 있다. 따라서 해당 정비자와 정비가 '○○P반'의 전문 정비자와 주요 정비내용일 가능성이 크다.

이처럼 워드 클라우드를 이용하면 주요 정비자와 정비내용을 직관적으로 확인할 수 있지만, 정비자의 전문성 확인을 위해 필요한 정비자 간 또는 '정비자-정비내용' 관계를 살펴볼 수 없다. 따라서 이를 위해 단어 중심의 SNA(워드 네트워크)가 필요하며, 출현빈도 상위 약 100개 단어를 대상으로 다음의 <그림 4>와 같이 시각화했다.

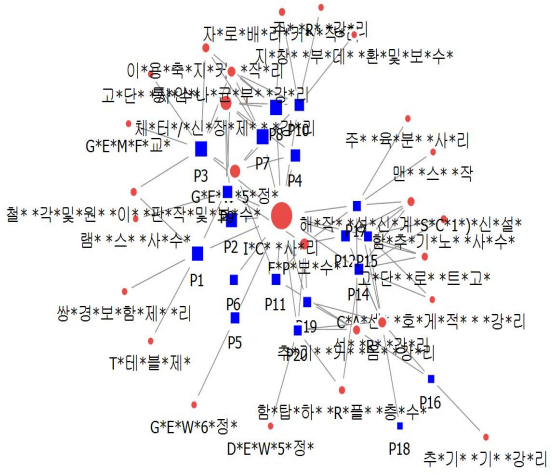

Fig. 4. Word Network(Maintenance detail)

여기서 청색 사각형은 정비자를 의미하고, 적색 원은 정비내용을 나타내며, 사각형과 원의 크기는 중심성의 높낮이(세기)를 의미한다. 즉, 단어 중심의 SNA 결과를 통해서 살펴보면 주요 정비자는 P1, P2, P3이고 주요 정비내용은 'I*C* *사*리'와'체*터*/*신*장*제* * *강*리'로 워드 클라우드와 유사하게 나타나고 있다. 또한 '정비자-정비내용' 간 관계의 전체적인 구조를 보면 정비자별로 주 정비 분야가 나누어져 있으며 특히, P16, P18은 노드 크기도 작고 주요 정비내용인 'I*C* *사*리'와 직접 연결링크가 없어 해당 정비에 참여하지 않았음을 확인할 수 있다.

4.2 Result of SNA

SNA 적용은 노드를 선택하고 노드 간 방향성과 가중치 값을 적용할 수 있다. 따라서 정비자의 정비능력이 숫자로 나타난다면 이 값을 가중치로 적용한 분석도 가능하다. '세부 정비내역'의 37개 속성 중 가중치 적용이 가능한 숫자 값을 가지는 속성에는 정비일자와 정비시간이 있으며 정비자의 정비 실시횟수도 중요하지만, 얼마나 오랫동안 정비했는지를 나타내는 정비시간 역시 고려해야 하므로 해당 속성을 가중치 값으로 적용하였다.

4.2.1 Analysis of Collaboration Relationship

일반적으로 해결하기 어려운 상황에 처했을 때 이를 믿고 맡길 수 있는 사람에게 의지하게 된다. 함정정비도 마찬가지로 일반적으로 해결하기 어려운 정비가 있으면 보다 실력있는 정비자의 도움을 받아 정비할 것이고, 따라서 정비자 간 협업이 높은 정비자일수록 어려울 때 해결사 역할을 하는 전문가일 가능성이 크다고 할 수 있다.

정비자 간의 협업 관계는 동일한 정비에 함께 참여했는지에 따라 나타나고 이는 정비관리번호를 통해 확인 가능하므로 정비자를 메인노드, 정비관리번호를 서브노드, 정비시간을 가중치로 적용하여 네트워크 분석을 진행하였다. 또한, 기본적으로 정비자와 정비관리번호는 서로 다른 집합이기에 2-Mode 네트워크 구성 후 1-Mode 네트워크로 변환하여 메인노드인 정비자 간의 협업 관계를 아래의 <그림 5>와 같이 확인하였다.

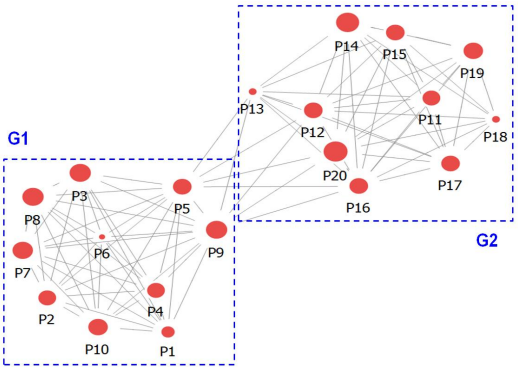

Fig. 5. Collaboration Relation Of Maintenance personnel

'○○P반'의 협업 관계를 살펴보면 적색 원은 모두 정비자를 나타내며 크게 2개 그룹(G1, G2)으로 나뉜다는 것을 알 수 있다. 실제 '○○P반'의 조직은 1반과 2반 총 2개의 반으로 구성되어 있으며 <그림 5>에서 정비자 P6, P13, P18은 노드의 크기가 작아 상대적으로 협업수준이 낮음을 알 수 있다. 또한, 정비자 P9, P20는 노드 크기가 크고 2개 반을 잇는 협업 관계에 있어 정비 전문가일 가능성이 있다.

Table 2. Rank of Collaboration Relation

Maintenance personnel	Degree of collaboration	Rank
P20(✓)	0.1191	1
P14	0.1171	2
P9(✓)	0.1121	3
P8	0.1111	4
P3	0.1110	5
P7	0.1082	6
P19	0.1069	7
P10	0.1054	8
P12	0.1047	9
P16	0.1046	10
P15	0.1045	11
P17	0.1039	12
P5	0.1029	13
P4	0.1015	14
P2	0.1006	15
P11	0.0994	16
P1	0.0861	17
P18(✓)	0.0698	18
P13(✓)	0.0685	19
P6(✓)	0.0627	20

이를 정비자 협업 정도 순위인 위의 <표 2>를 통해서 살펴보면 정비자 P6, P13, P18은 하위 18위~20위를 차지하여 협업 관계가 낮지만 P20, P9는 상위순위로 협업 관계가 높다.(본문 언급된 정비자는 표에 (✓) 표시함) 또한, 세부적인 협업 관계를 살펴보기 위해 '정비자-정비관리번호' 2-mode 네트워크에 대하여 응집성 분석을 하면 아래의 <그림 6>와 같이 나타나며 청색 사각형은 정비자, 적색 원은 정비관리번호를 의미한다.

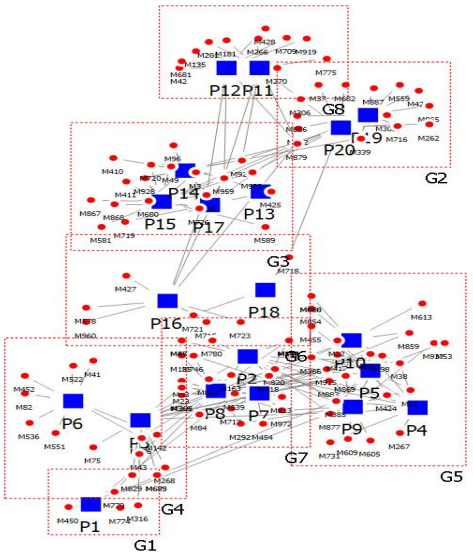

Fig. 6. Collaboration Relation Group(2-mode Cohesion)

세부 협업 관계는 총 9개 그룹(G1~G9)으로 구분되며, G1을 제외하고 각 그룹 G2~G9는 보통 2~4명의 정비자로 구성된다. 그리고 역시나 모두 자신의 소속 반(1, 2반) 내에서 상호 그룹을 이루어 협업 관계를 맺고 있음을 세부적으로 살펴볼 수 있다. 이러한 세부 협업 관계는 <그림 5>처럼 소속 반 내에서 상호 간에 협업 관계에 있지만, 그중에서도 자주 협업하는 관계를 더 상세하게 나타내며 일반적인 정비 시에는 <그림 6>과 같이 정비조를 이룬다는 것을 보여준다.

이처럼 협업 관계가 높은 정비자는 어려운 정비에 대해 해결사 역할을 하는 전문가일 가능성이 크다고 할 수 있겠다. 하지만 <그림 6>의 정비자 P1과 같이 특정 상황 및 정비 난이도 등에 따라 단독으로 정비를 실시하는 경우도 있어 정확도 높은 전문가 탐색을 위해서는 단순히 협업 관계뿐만 아니라 정비자가 해당 정비반에서 얼마나 중요한 정비에 자주 참여하였는지에 대한 것도 함께 고려해야 한다. 따라서 협업 관계만으로는 전문가를 판단하기에는 아직 부족하다고 할 수 있으며, 다음에서 알아볼 정비 전문분야 중심성 분석결과와 같이 종합적으로 판단해야 할 것이다.

4.2.2 Analysis of Centrality

일반적으로 중심이 되는 인물은 해당 조직 또는 반에서 중요한 업무를 담당한다. 정비 전문가는 해당 정비반에서 상대적으로 중요한 정비를 담당할 것이며, 정비의 개별 중요도는 전체 정비 중 해당 정비가 차지하는 비율을 통해서 살펴볼 수 있다. 총 10건의 정비 중 5건의 정비내용이 '정비A'이고 나머지 5건이 '정비B~정비F'라면 '정비A'가 해당 정비반의 주요 정비업무를 뜻하고 정비 전문가는 그 주요 정비에 많은 정비시간을 할애할 것이다. 따라서 정비자의 정비 전문분야를 살펴보고 위해서 정비자를 메인노드, 정비내용을 서브노드, 정비시간을 가중치로 적용하여 네트워크 분석을 진행하였다. 앞서 실시한 협업 관계와 같이 2-mode 네트워크 구성하여 분석 후 1-mode 네트워크로 변환하여 정비자 중심성 순위 등의 결과를 비교하였다. 또한, 전체 노드를 시각화하면 분석에 제한이 되어 2일(16시간) 이상 정비내용을 대상으로 하는 상위노드 약 100개를 <그림 7>과 같이 위세중심성 분석을 하였다.

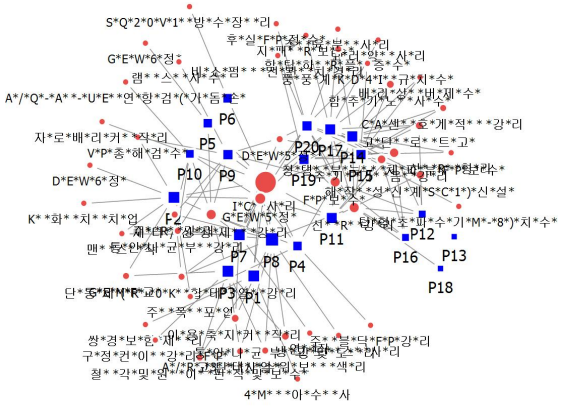

Fig. 7. Network of EigenVector Centrality
(Maintenance personnel-Maintenance detail)

<그림 7>에서 정비자는 P1, P7, P8 등이 정비내용은 'I*C**사*리'와 'G*E*W*5*정*'가 노드 크기가 커서 중심성이 높음을 알 수 있다. 이는 해당 '정비자-정비내용' 관계에서 해당 노드의 등장 빈도가 높으며 정비시간 면에서도 시간이 많이 할애 되었음을 뜻하고, 반면에 P13은 고립된 노드로 사전 설정한 조건에 따라 P13이 실시한 정비내용에 2일 이상(16시간) 정비실적이 없음을 나타낸다. 또한, 가운데를 기준으로 정비자들이 왼쪽은 1반, 오른쪽은 2반으로 나뉘어 있으며, 정비자별로 정비내용의 분야가 조금씩 다름을 통해 역시 함정정비가 정비반별, 개인별로 분업화되어 실시함을 확인할 수 있다. 이를 정비자의 중심성 연결정도 순위인 아래의 <표 3>를 통해서 다시 살펴보면 P8에 0.270으로 가장 순위가 높으며 P13은 0으로 가장 순위가 낮음을 알 수 있다.

Table 3. EigenVector Centr of Maintenance personnel

Maintenance personnel	EigenVector Centrality	Rank
P8(✓)	0.270	1
P1	0.216	2
P7	0.210	3
P3	0.209	4
P2	0.205	5
P14	0.178	6
P17	0.177	7
P11	0.172	8
P15	0.164	9
P20	0.161	10
P9	0.146	11
P19	0.142	12
P4	0.131	13
P5	0.121	14
P6	0.115	15
P10	0.095	16
P12	0.082	17
P16	0.062	18
P18	0.014	19
P13(✓)	0	20

다음으로 앞서 실시한 위세중심성 분석과 같은 조건(메인·서브

노드, 가중치)에서 페이지랭크 분석을 하면 <그림 7>와 전체적인 구조는 동일하나 일부 노드의 크기가 다르게 나타난다.(위세중심성과 페이지랭크 시각화 구조가 같아 그림은 생략함). 노드 크기의 차이점을 비교하기 위해서 위세중심성과 페이지랭크 중심성 값을 다음의 <표 4>와 같이 나타내면 계산 수식의 차이로 연결 정도와 정비자 순위가 P1의 경우 최대 -9까지 서로 다른 결과가 나온다.

세부적인 확인결과 이러한 차이는 정비자에 연결된 정비내용의 수에 따라 페이지랭크 값이 높아졌으며 즉, 정비자의 정비시간보다 우선 정비자가 참여한 정비내용의 수가 페이지랭크 값에 영향을 주었음을 알 수 있다. 물론, 정비자가 다양한 정비에 참여한 것도 중요하지만 전문성은 정비 경험을 기반으로 쌓이기 때문에 정비시간이 우선 고려되어야 한다. 따라서 비록 페이지랭크가 위세중심성을 문제점을 보완발전 시킨 분석방법이나 본 연구 목적에는 위세중심성이 더 적합하다고 할 수 있다.

Table 4. Comparison of EigenVector Cent & PageRank

Maintenance personnel	Degree of connection		Rank		
	EigenV	PageR	EigenV	PageR	Difference
P8	0.270	2.172	1	2	-1
P1(✓)	0.216	1.523	2	11	-9
P7	0.210	1.600	3	9	-6
P3	0.210	2.108	4	5	-1
P2	0.205	2.149	5	4	1
P14	0.178	1.671	6	7	-1
P17	0.178	2.168	7	3	4
P11	0.172	0.964	8	16	-8
P15	0.164	2.269	9	1	8
P20	0.161	1.862	10	6	4
P9	0.146	1.632	11	8	3
P19	0.142	1.162	12	13	-1
P4	0.131	1.235	13	12	1
P5	0.121	0.977	14	15	-1
P6	0.115	0.935	15	17	-2
P10	0.095	1.538	16	10	6
P12	0.082	1	17	14	3
P16	0.062	0.808	18	18	동일
P18	0.014	0.56	19	19	동일
P13	0	0	20	20	동일

4.2.3 Evaluation of SNA Results

전문가 탐색을 위해 앞서 알아본 정비자 간 협업 관계와 정비 전문분야 중심성 분석결과를 함께 살펴보면 아래의 <표 5>과 같다.

Table 5. Comparison of Collabo Rel & EigenV Centr

Maintenance personnel	Degree of connection		Rank		
	Collab	EV Cent	Collab	EV Cent	Difference
P8(✓)	0.1111	0.270	4	1	-3
P1	0.0861	0.216	17	2	15
P7(✓)	0.1082	0.210	6	3	3
P3(✓)	0.111	0.209	5	4	1
P2	0.1006	0.205	15	5	10
P14	0.1171	0.178	2	6	-4
P17	0.1039	0.177	12	7	5
P11	0.0994	0.172	16	8	8
P15	0.1045	0.164	11	9	2
P20	0.1191	0.161	1	10	-9
P9	0.1121	0.146	3	11	8
P19	0.1069	0.142	7	12	-5
P4	0.1015	0.131	14	13	1
P5	0.1029	0.121	13	14	-1
P6	0.0627	0.115	20	15	5
P10	0.1054	0.095	8	16	-8
P12	0.1046	0.082	9	17	-8
P16	0.1046	0.062	10	18	-8
P18	0.0698	0.014	18	19	-1
P13	0.0685	0	19	20	-1

정비자 P1의 경우에는 순위가 최대 +15로 차이가 커 협업 관계와 중심성 분석결과가 상호 연관성이 크지 않지만, 상·하위 2~3명을 보면 상호 유사한 결과를 보인다. 따라서 종합적으로 정비 전문가를 상위 2~3명에서 뽑는다고 할 때 정비자 P3, P7, P8이 '○○P반'의 전문가로 볼 수 있으며, 각 정비 분야를 세부적으로 살펴보면 다음의 <그림 8>과 같다.

서로 겹치는 정비내용을 제외하면 P3은 'D*E*W*5*정*', P7은 '선* *F* *강*리', P8은 'G*E*M*F*교*수*' 등의 정비에 경험이 있으므로 장비 작동원리 등에 대한 이해도가 높아 상대적으로 전문성을 가진다. 따라서 장비 고장 시 해당 장비에 정비 경험이 있는 정비자를 투입하면 신속한 장비복구에 효과가 있을 것이다. 물론, 다른 정비자들도 앞서 본 <그림 7>과 같이 상대적으로 자신이 많은 정비 경험이 있다면 비록 정비자의 순위가 낮더라도 해당 분야의 전문성을 가지고 있으며 <그림 8>은 '○○P반'에서 중요한 정비를 기준으로 많은 정비 경험을 가진 전문가를 탐색한 결과이다.

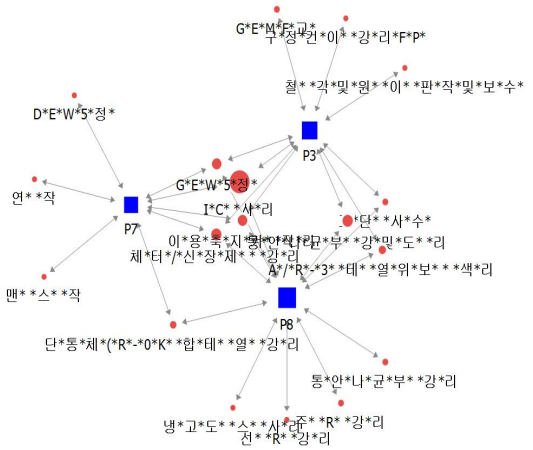

Fig. 8. Network of EigenVector Centrality
(Maintenance person of P3, P7, P8 detail)

다만, 중간순위 정비자들은 협업 관계와 정비 중심성 분석결과의 연관성이 낮은데 이는 중심성 또는 중요도가 높은 정비는 보통 정비반의 대부분 정비자가 정비에 참여하기에 협업 관계에 미치는 영향이 적고, 상대적으로 중심성이 낮은 정비는 일부 정비자들만 참여하므로 협업 관계에 더 큰 영향을 준다. 그런데 정비 전문가는 보통 정비반 내에서 중요도가 높은 정비에 주로 참여하기 때문에 경우에 따라 협업 관계와 정비 전문분야 중심성 결과의 값이 많은 차이를 보이기도 한다. 따라서 전문가 탐색에 있어 1차적으로 중심성 분석의 결과를 따르고 중심성 값

이 상호 근사한 경우 2차적으로 협업 관계를 통해서 종합적으로 판단해야 할 것이다.

4.3 Comparison of Analysis Results

본 연구에서 사용한 텍스트마이닝과 SNA의 가장 큰 차이점은 바로 노드의 선택 가능 유무로 텍스트마이닝은 노드를 자동 추출하여 노드 선택이 불가하나 SNA는 선택이 가능하다. 또한, 텍스트마이닝은 노드 추출에 있어 TF-IDF에 따라 빈도수만 고려되나 SNA는 빈도수와 별도의 가중치를 부여할 수 있다. 따라서 다음의 <표 6>과 같이 분석결과를 상호 비교하면 순위가 P17의 경우 최대 +8까지 차이 남을 알 수 있다.

물론 SNA가 노드의 등장 빈도 수와 가중치(정비시간)까지 고려되어 세부적으로 분석되었다고 할 수 있지만, 상·하위 2~3위 수준으로 대략적인 확인한다면 텍스트마이닝도 결과가 유사하게 나옴을 알 수 있다. 따라서 정비자 간의 협업 관계, 정비 전문분야 등 상세한 분석을 위해서는 SNA 방법이 적합하나 상·하위 2~3명의 대략적인 수준 정도만을 확인할 때는 SNA보다 간편하고 워드 클라우드를 통해 쉽게 이해할 수 있는 텍스트마이닝도 분석도 사용이 가능할 것으로 보인다.

Table 6. Comparison of Text Mining & SNA

Maintenance personnel	Degree of connection		Rank		
	Text M	SNA	Text M	SNA	Difference
P8	0.259	0.27	1	1	동일
P1	0.213	0.216	5	2	3
P7	0.246	0.21	2	3	-1
P3	0.243	0.209	3	4	-1
P2	0.234	0.205	4	5	-1
P14	0.141	0.178	10	6	4
P17(✓)	0.111	0.177	15	7	8
P11	0.137	0.172	11	8	3
P15	0.117	0.164	13	9	4
P20	0.111	0.161	14	10	4
P9	0.162	0.146	6	11	-5
P19	0.11	0.142	16	12	4
P4	0.158	0.131	7	13	-6
P5	0.144	0.121	9	14	-5
P6	0.104	0.115	17	15	2
P10	0.158	0.095	8	16	-8
P12	0.125	0.082	12	17	-5
P16	0.031	0.062	18	18	동일
P18	0.01	0.014	19	19	동일
P13	0	0	20	20	동일

현재까지 대부분의 전문가 식별 및 추천 등에 관한 연구는 텍스트마이닝 또는 SNA 중 1가지 분석방법만을 이용하였으나, 본 연구에서는 텍스트마이닝과 SNA 방법을 모두 이용해 결과를 비교 분석하여 해군 함정정비 전문가를 탐색하는 방법을 알아보았다. 또한, SNA에서는 위세중심성과 페이지랭크 중심성 결과를 비교하여 전문가 탐색을 위한 적합한 중심성 분석방법을 확인하였다. 이를 통해 식별된 정비 전문가는 함정이 군사작전 중 예기치 않는 장비 고장 시에 신속하게 해당 장비 전문가들을 파견하거나 원격 통신을 통한 정비지원 상담자로 활용할 수 있어 군의 정비능력 향상에 기여할 것이다. 그리고 민간 조직에서는 서론에서 소개한 것처럼 '저자-초록, 키워드' 또는 '질의-답변' 관계를 텍스트마이닝과 SNA을 통해서 해당 분야 전문가를 확인하고 문제와 관련된 조언 등을 통해서 신속한 문제 해결에 활용할 수 있을 것이다.

V. Conclusions

누가 어떠한 경험을 했다는 것이 데이터베이스 등에 기록되어 관리된다면 특정 상황에 유사한 경험을 가진 사람을 배치하여 신속하게 대응할 수 있을 것이며, 각종 복합적인 상황에서 전쟁을 수행해야 하는 군 조직에서는 더욱 경험이 중요한 요소라고 할 수 있다.

본 연구에서는 국방 분야에서는 거의 실시되지 않는 전문가 탐색에 관해 알아보았으며, 세부적으로 국방 장비정비정보체계의 2017년 해군 정비데이터 중 특정 정비반에 해당하는 일부(0,000건)를 데이터마이닝 기법을 적용하여 해군 함정정비 전문가를 파악하고자 하였다. 우선 텍스트마이닝을 통해 워드 클라우드 및 네트워크를 통해 대략적인 수준의 '○○P반'의 대표 정비자 및 정비내용을 알아보고 이후 SNA를 통해 '정비자-정비내용(가중치 : 정비시간)' 네트워크를 구성하여 세부적인 정비자 간 협업 관계, 정비 전문분야 및 분업화 등을 확인할 수 있었다. 특히, SNA에서는 2가지(위세중심성, 페이지랭크) 중심성 분석방법을 적용하고 이를 비교 분석하여 전문가 탐색결과 검증 및 연구 목적에 맞는 적합한 중심성 방법을 확인하였다.

하지만 본 연구는 해군 ○○창의 77개 정비반 중에서 1개 반만을 분석 대상으로 하였기에 아직 앞에서 알아본 분석방법의 정확성 여부를 확신할 수 없다. 왜냐하면, 정비반 조직특성에 따라 동일한 분석방법을 시도하더라도 정비데이터의 양이 적거나 정비자 인원수가 적으면 전문가 탐색이 제한될 수도 있기 때문이다. 또한, 1년이라는 제한적인 기간의 정비데이터를 기반으로 분석했기에, 향후 연구 방향으로는 분석결과의 정확성 향상을 위해 정비데이터 수집 기간을 늘리고 분석 대상을 특정 정비부대 정비팀(반) 전체로 확대하여 적어도 하나의 정비부대 전반에 걸친 연도별 시계열적 변화 확인 등 정비데이터에 대한 상세한 분석을 통해 함정정비 전문가를 확인해야 할 것이다. 또한, 현재 국방 장비정비정보체계에서는 정비데이터 현황을 보여주는 수준으로 현 체계를 통해서는 전문가 확인이 제한되므로 정비결과가 기록된 정비데이터를 이용하여 워드 클라우드나 네트워크 분석을 통해 시각화할 수 있도록 향후 국방 장비정비정보체계의 개선이 필요할 것으로 보인다.

REFERENCES

[1] Jeon. IB et.al, “An Expert Search Scheme for Improving Reliability in Social Network Environments” Journal of KIISE : Computing Practices and Letters 20(5), pp.311-315, May 2014.
[2] KIM. HS et.al, “Expert System” Jipmoondang, pp. 3-280. 1995.
[3] CHUN. MJ and HU. MS, “Knowledge Management System” Hangungsa, pp. 283-297. 2005.
[4] Kim. JG et.al, “Expert Exploration Using Social Network Analysis” Proceedings of The Korean Society of Computer Information Conference 27(1), pp.171-174, Jan 2019
[5] Linoff, Gordon S, Berry, Micahel J.(2011) Data mining techniques : for marketing, sales, and customer relationship management. Indianapolis, IN : Wiley.
[6] Lim. SS et.al, “A study on military organixational tasks analysis methology” Journal of the Korean Data And Information Science Sociaty 30, pp. 139-157, Jan 2019.
[7] KIM. YH and KIM. YJ, “Social Network Analysis. 4th edition.” Pakyoungsa, pp. 5-130. 2016.
[8] Ship’s Maintenance System, ROK Navy(2018)
[9] Cyram. NetMiner. 4.4. Seoul: Cyram Inc. 2018.
[10] Zafarani, et.al, “Social Media Mining” Cambridge Univ, pp. 75-81. 2014.

Authors

Jin-Gwang Kim received the B.S. degree in Computer Science from Korea Naval Academy in 2008 and He is a candidate for M.S. degree in Computer Science from Korea National Defense University. He is interested in knowledge discovery, Text Mining and Social Network Analysis.

Soung-woong Yoon received the B.S. degree in Engineering from Hanyang University in 1992 and M.S. degree in Computer Science and Engineering from Korea National Defense University (KNDU), Korea, in 2004, respectively. Mr. Yoon completed Ph.D course of the Department of Social Informatics at Kyoto University, Japan, in 2014. He is interested in knowledge discovery, information retrieval, big data and Social Network Analysis.

Sang-Hoon Lee received the B.S. degree in Electronic Engineering from Sung Kyun Kwan University, M.S. degree in Computer Science from Yon Sei University, and Ph.D. degree in Information Engineering from Kyoto University, Japan, in 1978, 1987 and 1997, respectively Dr. Lee joined the faculty of the Department of Computer Science at Korea National Defense University, in 2000. He is currently a Professor in the Department of Computer Science, Korea National Defense University. He is interested in DataBase, Big Data, Social Network Analysis, and Information search.