

데이터 마이닝 기반의 군사특기 분류 방법론 연구 (A Data-Mining-based Methodology for Military Occupational Specialty Assignment)

민규식, 정지원, 최인찬*

Abstract

In this paper, we propose a new data-mining-based methodology for military occupational specialty assignment. The proposed methodology consists of two phases, feature selection and man-power assignment. In the first phase, the k-means partitioning algorithm and the optimal variable weighting algorithm are used to determine attribute weights. We address limitations of the optimal variable weighting algorithm and suggest a quadratic programming model that can handle categorical variables and non-contributory trivial variables. In the second phase, we present an integer programming model to deal with a man-power assignment problem. In the model, constraints on demand-supply requirements and training capacity are considered. Moreover, the attribute weights obtained in the first phase for each specialty are used to measure dissimilarity. Results of a computational experiment using real-world data are provided along with some analysis.

* 고려대학교 공과대학원

1. 서 론

육군은 신병에 대한 군사특기 분류를 통하여 개인의 적성에 적합한 군사특기를 부여하고 이에 따라 보직을 배정한다. 따라서 적절한 군사특기 분류는 개인의 임무 충족도 뿐 아니라 군의 사기 및 전투력과 직결되어 있다. 최근, 저 연령층의 출산율 감소에 따른 징집가능 자원의 감소(통계청, 「장기 인구추계: 1990-2021」, 1991), 현역병에 대한 복무기간의 단축추세(국방일보, 2003.1.11) 등은 군사특기 분류의 중요성을 증가시키고 있다.

현재의 군사특기 분류 방법은 입영장정의 개인 자력과 신체/인성/지능 검사, 전공 관련 실기측정 등 속성별 점수를 근거로 하여 특기를 분류하고 있다. 현 특기분류 방법은 속성별 가중치를 사용하여 개인 점수를 산정함으로써 대외적인 투명성과 공정성을 어느 정도 유지할 수 있으나, 평가 항목의 선택과 가중치 산정에서 의사 결정자의 주관적 판단을 배제하기 어렵다. 본 논문은 이러한 문제점을 극복하기 위하여 과거의 경험적 데이터를 이용한 데이터 마이닝 기반의 군사특기 분류 방법론을 제시하는 것을 목적으로 한다.

데이터 마이닝은 축적된 데이터로부터 내재된 유용한 패턴이나 모형을 추출하는 일련의 과정이다 [8]. 데이터로부터 내재된 지식을 추출하기 위해서, 지식추론과 관련된 자동화 또는 반자동화된 다양한 기법들이 데이터 마이닝에 적용된다. 이러한 기법들 가운데 클러스터링(clustering)은 유사한 성향을 가진 몇 개의 클러스터(그룹)로 데이터를 구분하여 그룹간의 관계로부터 필요한 정보를 추출할 수 있도록 하는 자율 지식추론 기법이다[25]. 이와 같은

클러스터링 기법의 특성을 이용하여 특기별로 바람직한 적합집단을 산출하고 분류에 기여한 항목과 가중치를 산출해낼 수 있다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 인력 할당 문제와 관련된 기존연구를 살펴보고 3장에서는 속성 부분집합(subset) 기반의 클러스터링 기법을 이용한, 2단계로 구성된 데이터 마이닝 적용 방법론을 제시한다. 1단계는 속성선택 및 가중치 산정 문제로서, 기존의 최적 속성 가중치 산정 알고리즘의 문제점을 분석하고 이를 극복할 수 있는 2차 수리계획 모형(Quadratic Programming)을 제시한다. 2단계에서는 1단계에서 구해진 특기별 가중치를 이용하여 최적의 특기에 신병을 할당하는 인력 할당 문제를 현실적 제약이 고려된 정수 계획 모형(Integer Programming)으로 정식화한다. 4장에서는 실험을 통하여 제안된 방법론에 대한 적용 절차를 상세히 설명하며, 결론과 함께 향후 연구 방향을 5장에서 논의한다.

2. 기존연구

인력 할당 문제는 최적의 임무 수행을 위하여 제한된 가용 인력자원을 임무에 배분하는 문제로서 다양한 연구가 진행되어 왔다. 정해진 기간 내에 부여된 임무를 수행하기 위하여 작업자의 작업 순서와 일정을 결정하는 모형[5,6,7,9,17,18,19]로부터 작업 능력을 고려하여 개인의 작업량을 평균화하는 모형[3,4,12], 다양한 현실적 작업 조건을 고려하여 최소의 인력으로 최적의 임무 수행이 가능토록 작업자를 할당하는 모형[12,22,23]등 작업 순서 제약 및 부하 균형 목적의 존재 여부에 따라 인력 일정

계획 모형, 인력 부하 배분 모형 및 순수 인력 할당 모형으로 크게 분류할 수 있다.

단기간 내에 최상의 전투력을 갖추도록 교육하고 훈련시켜 자신의 적성에 맞는 직무를 수행하도록 해야 하는 군의 경우에 인력 할당 문제는 더욱 중요하다. 선진국의 경우 각 군(육,해,공,해병)별로 독특한 구비 조건 및 구조, 직무 패턴이 다른 상황을 반영하여 각각의 병력을 직무에 할당하는 최적화 네트워크 모형을 현실에 이미 적용하고 있다.

Klingman 등[11]은 네트워크 모형을 이용하여 다른 요구 조건을 가진 우선순위가 있는 직무에 해병 병사를 할당하는 모형을 제안하였다. 이 모형은 인력 계획과 인력 할당을 통합하는데 소요되는 비용을 최소화함을 목적으로 한다. Baumgarten[2]은 필수 기간을 가진 일련의 직무로 구성된 적합한 경력 경로를 창안하고 이러한 경력 경로와 각 직무가 요구하는 개인 필수 조건에 근거한 정수계획법을 사용하여 경력 경로를 거쳐야 할 해병 장교의 수를 결정하는 모형을 제안하였다. Sweeny[20]는 네트워크 모형을 이용하여 해병 장교를 직무에 할당하는 순수 할당 모형을 제안하였다.

Sweetser 등[21]은 계획된 육군 요구 사항과 부대 전투력에 기초한 직무에 병사를 할당하고 재분배하는 통계, 시뮬레이션, 최적화 통합 모형을 제안하였다. Zuidweg 등[24]은 전투 중 발생할 수 있는 적 공격으로부터 함정 근무자들의 생존율과 함정 생존율을 통계적으로 분석하고 그 통계자료에 의해 최적화된 근무 인원을 전투함에 할당하는 수리적 모형을 제안하였다. Liang과 Thompson[13]은 단일 가중치를 가진 목적함수를 이용하여 해군 수병을 우선순위에 따른 직무에 할당하는 대규모 인력 할

당 모형을 제안하였다. 이 모형은 개인별 자격과 직무별 요구 자격 조건을 고려하여 유사한 기술을 가진 개인들을 할당한다. Liang과 Buclatin [14]은 요구되는 새로운 기술과 능력을 보유 또는 교육 이수한 인원을 새로운 직무에 할당하는 모형을 제안하였다. 이 모형은 네트워크 모형을 선형계획 모형으로 전환하고 비용 최소화 문제를 효용 최대화 문제로 전환한 후, 개인의 능력을 최대도 발휘할 수 있는 효용 척도를 최대화한다.

기존의 군사특기 분류를 위한 할당모형에서 다목적함수 모형은 문제의 표현이 비교적 명확하지만, 목적함수간의 우선순위를 임의로 가정한다. 반면에 Liang 등[13,14]의 단일목적함수는 상대적으로 풀기 쉬운 모형을 제공할 수 있지만, 속성별 효용이나 가중치 산정에 의사 결정자의 주관이 개입될 수 있는 문제점이 있다.

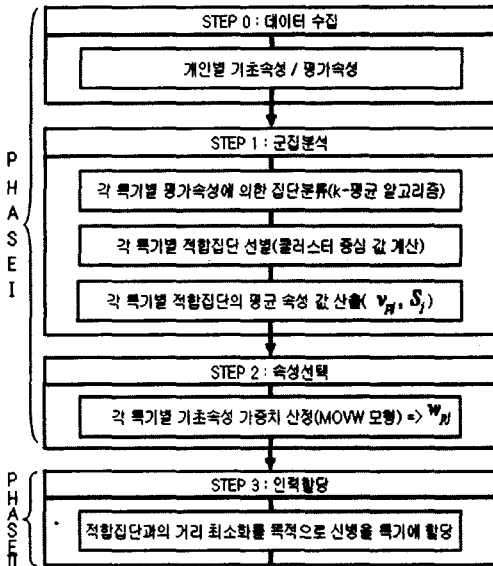
3. 데이터 마이닝 기법을 이용한 군사특기 분류 방법론

3.1 육군 신병 군사특기 분류 방법론

신병의 군사특기 분류를 위하여 본 논문에서 제시하는 방법론의 1단계에서는 각 특기별로 이미 특기를 부여받아 자대에서 임무를 수행중인 인원들의 개인에 관한 축적된 자료를 이용하여 집단을 분류한다. k-평균 알고리즘을 통해 해당 특기에 가장 적합하게 임무를 수행하고 있는 적합집단을 선별하고, OVW(Optimal Variable Weighting) 모형[16]을 수정 적용하여 얻게 되는 군집분석의 결과로부터 속성별 가중치를 산출한다.

2단계 할당문제에서는 1단계에서 구해진 속성별

가중치를 고려하여, 새로 특기를 부여할 신병의 개인별 속성 값이 할당 대상이 되는 각 특기별 적합 집단의 평균 속성과 가장 유사한 특기에 신병을 할당한다. 이러한 접근방식은 각 속성에 주관적 가중치를 부여하여 점수로 합산하는 기존 방식과는 구분된다. 제시되는 방법론의 절차를 세부 단계별로 나타내면 <그림 1>과 같다.



<그림 1> 방법론 절차

제시된 방법론은 실제 데이터를 기반으로 적합 집단을 분류하고 기여도가 큰 속성을 산출하기 때문에 의사 결정자의 주관적 판단을 배제하고 객관적이고 공정한 판단 기준을 제공한다. 따라서 보다 의미 있는 결과를 도출하기 위하여 STEP 0의 데이터 수집 단계는 매우 중요하다. 즉, 의미 있는 자료가 배제되거나 오류가 과다하게 포함되어 있는 경우 적절한 결과를 기대하기 어렵다. 이러한 데이터 준비와 정리 단계는 데이터 웨어하우징과 클린

징 분야에 관련된 것으로 본 논문의 취지에 적합하지 않으므로 상세한 언급을 피한다.

STEP 1에서는 각 특기에 기 할당된 인원을 대상으로 평가속성을 이용한 군집분석을 통해 집단을 분류한다. 이때 주어진 k 값에 따라 우수집단 및 비우수집단과 몇 개의 중간층 집단으로 분류되며, 산출된 집단에서 가장 우수한 평균성적을 갖는 집단을 적합집단으로 선정한다. STEP 1의 적합집단 분류를 위해서는 각 특기에 기 할당된 인원의 평가속성만이 사용된다. 평가속성은 입영 신병이 주특기를 부여 받은 후 각종 교육기관과 자대에서의 임무 수행능력을 평가받은 결과를 의미한다. 그러나 신병의 할당을 위하여 신병이 입영 전까지 개인이 가지는 고유 기록으로서의 기초속성에 대한 속성별 가중치가 필요하므로, STEP 2의 수정된 OVW (Modified OVW ; MOVW) 모형에서는 각 특기에 기 할당된 인원의 기초속성을 사용하여 속성별 가중치를 구하게 된다. STEP 3에서는 할당하고자 하는 신병들에 대한 기초속성의 속성 값과 STEP 2에서 구해진 가중치를 이용하여 각 특기별 적합집단의 평균 특성에 가장 가까운 특기에 신병을 할당하게 된다.

3.2 데이터 수집

본 논문에서는 군집분석에 이용되는 데이터의 속성들을 평가속성과 기초속성으로 분류한다. 기초속성은 신체상태, 지적상태, 가정환경, 성향 등을 포함할 수 있으며, 개인별 임무 수행능력을 나타내는 평가속성은 교육기관 또는 자대 배치 후의 모든 평가 관련 속성들을 이용할 수 있다.

본 방법론에서는 MOVW 모형에 의하여 평가와

할당에 필요한 속성의 중요도를 산정하므로, 필요한 속성이나 속성의 중요도에 대한 사전 판단이나 제약을 두지 않는다. 즉, 기초속성과 평가속성에 대하여 가용한 모든 속성 자료를 이용할 수 있으며, 이들 속성의 의미와 중요도는 가중치 산정을 통해 판단되고 불필요한 속성들은 제거된다. 그러므로 가능한 모든 자료를 분석에 포함시켜 의미 있는 속성이 사전에 차단되지 않도록 하여야 한다.

3.3 k-평균을 이용한 군집분석

k-평균 알고리즘은 비계층적 클러스터링의 대표적인 기법이다. 가상의 k개 초기 클러스터 중심을 이용하여 가장 근접한 중심에 데이터를 할당하며, 중심값의 변동이 없을 때까지 가상 중심을 할당된 객체의 평균값으로 이동시키면서 반복적으로 데이터를 할당한다[15].

k-평균 알고리즘은 주어진 데이터 셋(set)을 동일한 k개의 클러스터로 구분하는 것으로 집단을 특정한 목적 속성을 대상으로 구분하는데 적합하다. 또한, 상대적으로 대규모 데이터 셋을 처리하는데 효과적[1,10]인 것으로 알려져 있으며, 수치형 데이터만을 대상으로 한다. 우수 또는 적합집단을 산출하기 위해 사용되는 평가속성은 일반적으로 개인의 수행능력을 수치화한 것으로서 주로 숫자형 데이터들로 구성된다. 간혹 범주형 데이터가 사용되는 경우도 있지만 순차적으로 정렬이 가능하므로 쉽게 수치형 데이터로 전환할 수 있다.

k-평균 알고리즘에서는 평가속성들만을 이용하여 과거 해당 주특기에 할당된 인원을 대상으로 우수 또는 적합집단을 분류해낸다. 도출된 적합집단의 평균값과 표준편차를 계산하여 3.5절의 인력할

당 문제에서 각 주특기별 적합집단의 평균특성 산정과 할당할 신병과의 비유사성(dissimilarity) 계산에 사용한다.

3.4 MOVW 모형을 이용한 속성선택

k-평균 알고리즘에 의해 구해진 클러스터링 결과로부터 클러스터링에 기여한 의미 있는 속성에 대한 가중치를 산출하기 위하여 Makarenkov 등 [16]은 OVW 모형을 제안하였다. 그러나 OVW 모형에서는 범주형 데이터는 연속성을 갖는 수치형으로 전환되어야만 적용이 가능 하다. 또한, 범주형 데이터를 수치형 값으로 변환 하는 과정에서 범주 영역이 순차적이지 않은 경우에는 의미 있는 결과를 생성하기 어렵다[10].

수정된 OVW 모형에서는 범주형 데이터 적용의 문제점을 보완하기 위하여 k-평균 개념을 범주형 데이터 영역으로 확장한 k-표준(k-Prototype) 알고리즘[10]의 거리 산정 방식을 이용한다.

우선, 범주형 데이터 속성들을 유사한 성격을 갖는 몇 개의 카테고리로 분리하고 각 카테고리가 다수의 부속성(sub-attribute)을 포함하는 계층적 형태로 표현한다. 이러한 계층 구조에서 각 카테고리별로 두 객체간 동일 값을 갖는 부속성의 수에 비례하는 거리를 산정할 수 있다. 데이터가 연속형 수치 값을 갖는 경우에는 다음 식 (1)과 같이 객체간의 유클리디안 거리를 이용한다.

$$d_{ij} = \left[\sum_{p=1}^m (y_{ip} - y_{jp})^2 \right]^{\frac{1}{2}} \quad (1)$$

여기서 i, j 는 클러스터 내 각 객체를, p 는 속성을

표기한다.

범주형 데이터에서는 식 (2)와 같이 거리를 산정한다. 식 (2)는 두 객체 i, j 간의 카테고리 p 에서의 범주형 데이터에 대한 거리를 산정하는 식이며, 식 (3)은 카테고리 p 에 속한 q 번째 부속성에 대한 거리를 구하는 식이다. 식 (2)와 (3)은 카테고리 p 에 속한 q 개의 부속성에서 객체 i 와 객체 j 가 다른 값을 갖는 부속성의 수를 구하고, 카테고리 p 의 전체 부속성의 수로 나눈 값을 거리로 사용한다.

$$d_{ij}^c = \left[\sum_{q=1}^r \delta(y_{ipk}^q, y_{jpk}^q) \right] / r_p \quad (2)$$

$$\delta(y_{ipk}^q, y_{jpk}^q) = \begin{cases} 0 & (y_{ipk}^q = y_{jpk}^q) \\ 1 & (y_{ipk}^q \neq y_{jpk}^q) \end{cases} \quad (3)$$

수치형 데이터와 범주형 데이터가 혼합되어 있는 경우 아래 식 (4)와 같이 거리를 계산할 수 있다.

$$d_{ij} = \left[\sum_{p=1}^r (y_{ip} - y_{jp})^2 + \sum_{p=r+1}^m d_{ijp}^c (y_{ip}, y_{jp})^2 \right]^{\frac{1}{2}} \quad (4)$$

Makarenkov 등[16]의 OVW 모형은 범주형 데이터를 사용할 수 없는 제한점 외에도 자명해에 빠지기 쉽다는 점과 모든 데이터 값이 동일한 의미 없는 속성을 제거하지 못한다는 문제점이 있다. 아래 OVW의 수리적 모형은 n 개 객체와 m 개 속성을 갖는 2 차원 배열 데이터 행렬 Y 가 주어지면, 각 객체간 거리를 구한 행렬 $D = [d_{ij}]$ 에 의해 각 속성에 대한 가중치 $w = \{w_1, w_2, \dots, w_m\}$ 를 산출한다.

$$\min L_p = \sum_{k=1}^K \left[\sum_{i,j=1}^{n_k} d_{ij}^2 \right] / n_k$$

$$\text{s.t.} \quad \sum_{p=1}^m w_p = 1$$

$$w_1, w_2, \dots, w_m \geq 0$$

$$\text{where } d_{ij} = \left[\sum_{p=1}^m w_p (y_{ip} - y_{jp})^2 \right]^{\frac{1}{2}}$$

[기호 설명]

d_{ij} : 클러스터 k 내 객체 i 와 j 간의 거리

n_k : 클러스터 k 내 객체 수

w_p : 속성 p 의 가중치

y_{ip} : 객체 i 가 가진 속성 p 의 좌표값

목적함수 L_p 는 클러스터 내부의 객체간 거리를 최소화하는 가중치를 구한다. OVW 모형의 목적식을 재정식화하면 다음과 같다.

$$\begin{aligned} \sum_{k=1}^K \left[\sum_{i,j=1}^{n_k} d_{ij}^2 \right] / n_k &= \sum_{k=1}^K \left[\sum_{i,j=1}^{n_k} \sum_{p=1}^m w_p (y_{ip} - y_{jp})^2 \right] / n_k \\ &= \sum_{p=1}^m w_p \left[\sum_{k=1}^K \left\{ \sum_{i,j=1}^{n_k} (y_{ip} - y_{jp})^2 \right\} / n_k \right] \\ &= \sum_{p=1}^m A_p w_p \end{aligned}$$

여기서, A_p 는 속성 p 에 대한 클러스터별 객체간 거리의 제곱합의 총합을 나타내며, 주어진 행렬 Y 로부터 계산되는 상수이다. 따라서 목적함수 L_p 는 결정변수의 합이 1인 단순한 선형문제로 귀결되고, A_p 가 가장 작은 속성의 가중치 w_p 는 1, 나머지는 0이 되는 자명해를 갖게 된다. 이것은 객체간

거리가 가장 작은 속성만을 선택함을 의미한다. 특히 모든 객체간 거리의 합이 0, 즉 A_p 가 0이 되는 속성이 여러 개 존재하는 경우에는 무한한 조합의 해가 존재하여 무의미한 가중치를 산출하게 된다.

OVW 모형의 이러한 문제를 해결하기 위해 결정변수 w_p 의 2차식(quadratic term)을 이용하고, 모든 객체간 거리의 합이 0인 속성은 제거되도록 제약식을 추가함으로써 자명해의 발생을 최소화할 수 있다. 또한, 범주형 데이터를 포함하는 경우에도 속성 가중치를 산출할 수 있도록 k-표준 알고리즘의 거리 산정 식을 이용하여 OVW 모형을 수정한다. 본 문서에서 제시하는 MOVW 모형은 다음과 같다. 여기서 M 은 임의의 큰 수를 나타낸다.

<MOVW: Modified Optimal Variable Weighting>

$$\min \sum_{k=1}^K [\sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} d_{ijk}^2] / n_k \quad (5)$$

$$\text{s.t.} \quad \sum_{p=1}^m w_p = 1 \quad (6)$$

$$M \cdot \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} d_{ijp}^2 - w_p \geq 0 \text{ for all } p \quad (7)$$

$$w_1, w_2, \dots, w_m \geq 0 \quad (8)$$

$$\text{where } d_{ijk} = [\sum_{p=1}^r \{w_p (y_{ipk} - y_{jp k})\}^2] \quad (9)$$

$$+ \sum_{p=r+1}^m \{w_p \cdot d_{ijp}^c (y_{ipk}, y_{jp k})\}^2]^{\frac{1}{2}}$$

[기호 설명]

d_{ijk} : 클러스터 k 내 객체 i 와 j 간의 거리

d_{ijp} : 속성 p 에 대한 객체 i 와 j 간의 거리

n : 전체 객체 수

y_{ipk} : 클러스터 k 내 객체 i 가 가진 속성 p 의 좌표값

식 (6)과 (8)은 속성 가중치가 양의 값을 가지며, 가중치의 합이 1이 되도록 제한한다. 식 (7)은 속성 p 에 대해 모든 데이터 값이 동일할 때 무의미해지는 해당 속성을 제거한다. 식 (9)는 식 (4)의 거리 산정 식에 가중치를 반영한 것으로, OVW 모형과는 달리 속성 가중치와 객체간 거리를 곱한 후 제곱함으로써 아래 식 (10)과 같이 속성 가중치를 2차 수리계획 모형에 의하여 산출되도록 한다.

목적식 (5)는 식 (9)에 의해 다음과 같이 재정식화 될 수 있다.

$$\begin{aligned} & \sum_{k=1}^K [\sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} d_{ijk}^2] / n_k \\ &= \sum_{p=1}^r w_p^2 [\sum_{k=1}^K \{ \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} (y_{ipk} - y_{jp k})^2 \} / n_k] \\ & \quad + \sum_{p=r+1}^m w_p^2 [\sum_{k=1}^K \{ \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} d_{ijp}^c (y_{ipk}, y_{jp k})^2 \} / n_k] \\ &= \sum_{p=1}^m A_p w_p^2 \quad (10) \end{aligned}$$

3.5 군사특기 할당 모형

할당 모형에서는 속성선택 단계를 통해 도출된 속성 가중치와 적합집단의 평균 특성을 이용하여 각 특기별 적합집단의 중심 값과 신병의 속성 값과의 거리를 구한 후 가장 근접한 적합집단의 특기에 신병을 할당한다. 본 문서에서는 군사특기 할당 모형을 다음과 같이 제안한다.

$$\min \sum_{i=1}^m \sum_{j=1}^n d_{ij}^2 \cdot x_{ij} \quad (11)$$

$$\text{s.t.} \quad \sum_{j=1}^n x_{ij} = 1 \quad \text{for all } i \quad (12)$$

$$\sum_{i=1}^m x_{ij} \leq R_j \quad \text{for all } j \quad (13)$$

$$\sum_{i=1}^m x_{ij} \geq \left(m \times R_j / \sum_{j=1}^n R_j \right) \quad \text{for all } j \quad (14)$$

$$\sum_{i=1}^m \sum_{j=1}^n a_{lj} \cdot x_{ij} \leq C_l \quad \text{for all } l \quad (15)$$

$$x_{ij} \in \{0, 1\}$$

$$\text{where } d_{ij} = \left[\sum_{p=1}^r \{w_{pj} (A n_{ipj} - v_{pj}) / S_j\}^2 \right. \quad (16)$$

$$\left. + \sum_{p=r+1}^m \{w_{pj} \cdot d_{ip}^c (A n_{ipj} - v_{pj}) / S_j\}^2 \right]^{\frac{1}{2}}$$

[기호 설명]

x_{ij} : 인원 i 가 특기 j 에 할당되면 1, 아니면 0

d_{ij} : 인원 i 와 특기 j 의 적합집단과의 거리

R_j : 특기 j 의 적정 총원소요

a_{lj} : 병과학교 l 에서 특기 j 를 교육하면 1, 아니면 0

C_l : 병과학교 l 의 교육 능력(장비, 교실, 교관, 조교, 숙식 고려)

w_{pj} : 특기 j 의 적합집단의 속성 p 의 가중치

$A n_{ipj}$: 인원 i 가 특기 j 에 할당되었을 때 속성 p 의 좌표값

v_{pj} : 특기 j 의 적합집단에 속하는 데이터들의 속성 p 의 중심점(평균) 좌표값

S_j : 특기 j 의 적합집단 데이터 표준 편차

식 (11)은 목적식으로 할당 인원과 적합집단과의 거리로 표현되는 비유사성의 합을 최소화하는 것을 목적으로 한다. 식 (12)는 각 개인은 반드시 1개의 특기에 할당되어야 함을 나타내며, 식 (13)은 각 특기별로 할당되는 총 인원은 해당 특기의 적정 총원소요를 초과할 수 없음을 나타낸다. 식 (14)는 수급 인력이 부족한 경우 인력의 공정한 배분을 위하여 최소 총원소요 수준 이상이 할당될 수 있도록 제한하고 있으며, 식 (15)는 각 특기에 할당되는 최대 인원은 각 병과 학교의 교육 능력 범위를 초과할 수 없음을 의미한다.

식 (16)은 할당할 인원 i 와 특기 j 간의 거리를 나타내는 것으로, 속성선택에서 구한 속성별 가중치를 반영하여, 특기별 적합집단의 중심 값까지의 거리를 해당 적합집단의 표준편차로 나눈 값이다. 거리 계산에서 표준편차를 이용함으로써 각 특기별 속성별 적합집단의 크기를 고려하여 비유사성을 계산할 수 있도록 한다.

4. 적용 방안 및 실험

위에서 제시한 모형을 현실에 적용할 경우에 사용할 데이터는 이미 특기를 부여받아 자대에서 임무를 수행중인 인원과 새로 입영하는 인원을 대상으로 수집하여야 한다. 또한, 제시된 모형의 적절성을 평가하기 위해서는 기초속성과 평가속성에 대한 모든 데이터를 가진 집단, 즉 이미 특기에 할당된 집단에 대한 데이터가 필요하다. 현실적으로 같은 시기에 동일한 특기를 부여받은 인원은 다양한 부대에 배치되어 있으며 동일한 방식으로 데이터를 관리하지 못하고 있는 문제점 등으로 인하여 데이

터 수집이 제한적이다. 본 연구에서는 수집이 가능하면서 요구되는 평가속성에 대하여 동일한 평가 기준을 가진 데이터만을 실험에 이용하였다.

임무가 서로 상이하어 특기별로 요구되는 속성에 대한 분석이 가능하도록 비교적 구분이 쉬운 7개 특기를 선정하였다. 선정된 특기는 육군 훈련소 후반기 교육이 필요한 2개 특기와 3개 병과학교에서 보충 교육이 필요한 5개 특기로 구성된다. 실험은 선정된 7개 특기에서 2003년에 육군 훈련소를 수료한 인원을 대상으로 하였다. 수집된 데이터의 특기별 인원수와 육군 훈련소 교육 종료 시기를 고려하여 군집분석 대상과 특기할당 대상으로 구분하였다. 즉, <표 1>에서와 같이 속성 가중치를 산출하기 위해 군집분석에 이용할 군집분석 대상과 임명 대상 신병을 대신하여 특기를 할당받게 될 특기할당 대상으로 구분한다.

<표 1> 특기별 실험 대상 인원

특기번호	특기명	군집분석 대상	특기할당 대상
1114	60M 박격포	503	166
1127	고속 유탄 기관총	145	68
1712	유선 시설 운용	132	57
1732	중계/반송기 운용	292	137
2421	무선 장비 수리	129	68
2511	경차량 수리	115	66
4111	일반 의무	195	57

기초속성은 수치형과 범주형 데이터를 포함하여 <표 2>에 명시되어 있으며, 수치형 중에서 18~21번 속성은 기술검사로서 개인이 각 특기별로 요구되는 기술검사에서 취득한 점수를 나타낸다 기초속성은 나이, 신장, 체중, 시력, 혈액형, 결합 사항, 신체등급 등의 신체상태와 IQ, 학력, 학과, 경력, 자

격, 직업 등 지적상태, 부모, 재산, 거주지 등 가정 환경, 종교, 인성검사, 입영 구분-지원 여부 등 개인적 성향 등이 있으나 이중 수치형과 범주형 데이터의 특징을 적절하게 표현하는 속성만을 선정하였으며 <표 2>는 기초속성 항목이다.

<표 2> 개인별 기초속성 항목

구분	번호	속성	부속성	구분	번호	속성	부속성			
수치형	1	입영구분		범주형	13	신체결합	운동			
	2	재산상태					신체결합			
	3	학력					허약			
	4	신장					시각			
	5	체중					청각			
	6	신체등급					17	혈액형	A	
	7	시력(좌)							B	
	8	시력(우)					18	기술	O	
	9	IQ							19	학력
	10	인성검사							20	자격
범주형	11	부모	양친	15	16	청각	안경			
			편부				시각장애			
			편모				난청			
			계부				언어장애			
			계모				A			
			양자				B			
			고아				O			
	12	신체결합	용모	흉터	18	기술	검사	AB		
				부정교합				19	학력	
				문신				20	자격	
			흉위이상	21			경력			
			O다리				실기			

개인별 임무 수행능력을 나타내는 평가속성은 우수한 집단과 비우수한 집단을 구분할 수 있는 기준이 되는 속성으로서 교육기관 수료성적, 자대 병공통과제 성적, 내무생활 상태, 상·하·동급자 평가, 해당 특기 임무 수행상태 등 다양하고 객관적인 평가 관련 항목들이 사용될 수 있다. 그러나 자료 수집의 제약으로 인해 각 특기별로 보충 교육기간의 평가 점수만을 사용하였으며, <표 3>에 정리되어 있다.

<그림 2>는 적용 절차에 대한 진행과 진행과정

에서 생성된 데이터의 출력과 입력을 정리한 것이다.

<표 3> 특기별 평가속성 항목

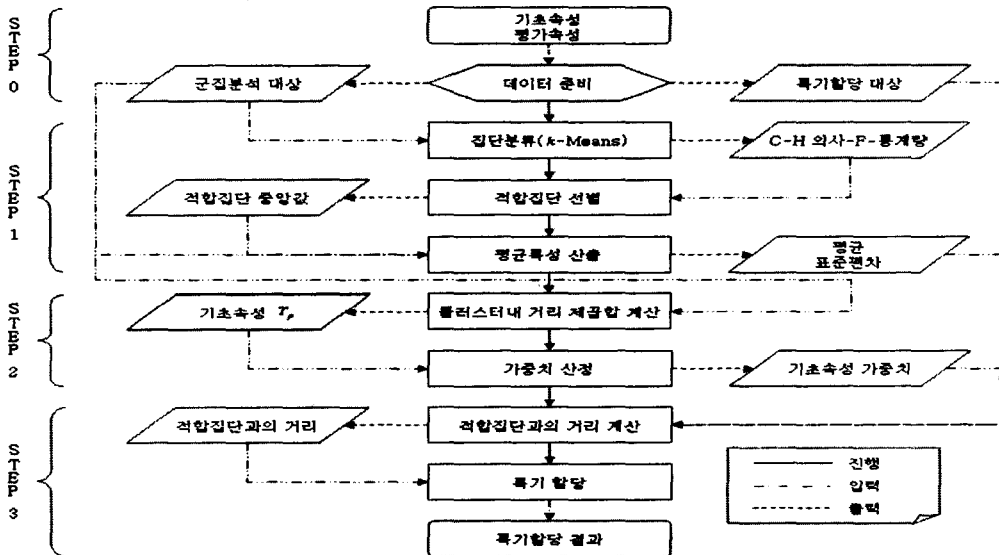
특기	1114	1127	2511	4111
평가 내용	종합	종합	요약	의무부대
			실습	야전응급 응급의학
특기	1712	1732	2421	의무군수
평가 내용	고정다중	고정다중	전술통신	해부의학
	내무평가	내무평가	FM장비	환자수송
	보안	보안	AM장비	환자간호
	정신교육	전술다중	최종평가	원무행정
	체육	정신교육	훈육평가	규정방침
	통신선로	체육		정신전력
		초단파통신		내무점수
		통신공학		
		통신운용		

먼저 적합집단 분류를 위해 군집분석 대상 인원의 특기별 평가속성을 기준으로 k-평균 클러스터링을 수행하여 클러스터 라벨을 결정하였다. 클러스터링은 Legendre[26]의 프로그램을 이용하였으며

데이터 값은 정규화(generalization)하였다.

분류된 클러스터의 중앙값이 가장 높은 클러스터를 적합집단으로 선정하였다. 중앙값은 각 클러스터에 속한 모든 객체의 속성값 평균이며, 중앙값이 가장 높은 집단은 해당 집단의 평균 성적이 가장 우수한 집단을 의미한다. 이는 평가속성은 주특기 부여 후의 개인별 평가 성적으로서 보다 큰 값이 우수하다는 기본적인 가정을 전제로 한다. 선정된 각 특기별 적합집단에서 기초속성에 대한 속성별 평균(v_{pj})과 표준편차(S_j)를 산출하여 마지막 인력할당 단계에서 사용될 수 있도록 하였다.

결정된 클러스터 라벨을 기준으로 MOVW 모형을 이용하여 특기별 기초속성 가중치를 산출하였다. 가중치는 정규화한 데이터를 이용하여 3.2절의 식 (10)에서 도출한 A_j 를 계산하여 2차 수리모형으로 전환하고 LINGO 8.0을 이용하여 최적해를 산출하였다.



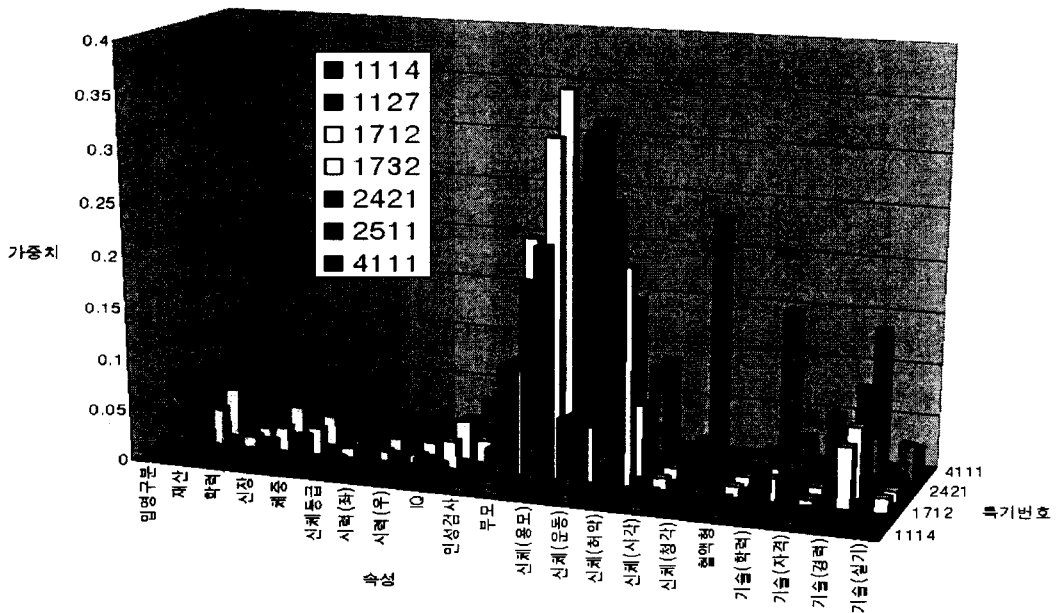
<그림 2> 적용 절차 흐름도

인력 할당을 위해 군집분석 결과와 특기할당 대상 데이터를 이용하여 3.3절 식 (16)의 d_{ij} 를 산출하였다. 식 (16)으로부터 산출한 d_{ij} 를 OPL Studio 3.5에 이용하여 인력 할당 모형에 의해 특기할당 대상 인원을 각 주특기에 할당하였다.

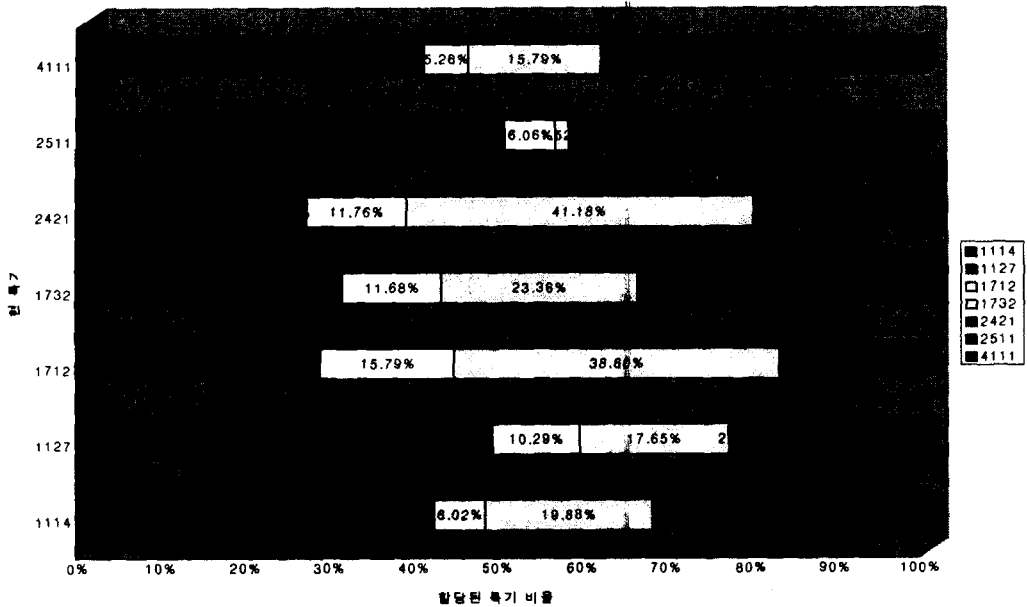
<그림 3>은 MOVW 모형에 의해 구해진 특기별 기초속성별 가중치를 막대그래프로 표현한 것이다. 색상별로 표시된 각 특기에 따라 속성별 가중치 값이 다양하게 나타나 있으며, 이는 각 특기별로 상이한 속성 가중치를 요구하고 있음을 나타낸다. 특기번호 2421을 제외한 각 특기의 신체결함(청각) 속성은 모든 값이 동일하여 분류에 기여하지 못하는 속성으로서 제약식 (7)에 의해 제거됨을 알 수 있다.

<그림 4>는 제안된 방법론에 의한 특기 할당 결과와 기존에 부여받은 현재의 특기와의 차이를

비율로 나타낸 것이다. 세로축은 현재의 특기이며 각 특기의 인원 중에서 새로운 특기에 할당된 인원의 비율이 가로축에 표시되어 있다. 수집된 데이터의 수가 적고, 평가속성에 대한 다양하고 정확한 데이터가 부족하여 실험결과에 대한 비교분석은 제한적 일수 밖에 없으나, <그림 4>에서 보는 바와 같이 7개 특기에 한정된 인원을 할당할 경우, 기존의 할당된 특기와는 상이한 결과를 볼 수 있다. 이러한 기존 방식과의 차이는 속성별 가중치의 변화에 기인하는 것으로 가중치 변화에 따른 할당 결과의 변화 폭이 매우 큼을 알 수 있다. 그러나 할당 결과의 비교, 평가를 위해서는 이미 근무 중인 인원의 변경될 특기에 대한 재평가가 요구되지만, 실제로 이미 근무 중인 인원을 대상으로 변경된 특기에서의 임무 수행능력을 재평가하는 것은 현실적으로 불가능하다.



<그림 3> 특기별 속성별 가중치 비교



<그림 4> 특기 할당 결과

과거의 경험적 데이터로부터 우수한 근무 평가를 갖는 적합집단을 추출하고, 이러한 적합집단과 유사한 패턴의 인력을 우선 배치한다는 절차상의 개념이 인간의 인지에 의한 판단과정과 유사할 뿐 아니라, 일부 전문가 집단에 의한 임의의 가중치 부과 방식에 비하여 데이터에 근거한 보다 합리적인 방법론을 제안한 것에 그 의의가 있다. 그러나 결과에 대한 정확도 향상을 위하여 의미 있는 데이터의 확보를 위한 적절한 방안이 강구되어야 할 것으로 판단된다.

5. 결 론

본 논문에서는 단일목적함수를 이용하여 상대적

으로 풀기 쉬우면서도 현실적으로 필요한 제약을 적절히 반영할 수 있는 모형과 이를 풀기 위한 방법론을 제시하였다. 이 모형은 기존 속성선택 기법의 문제점을 보완한 MOVW를 이용하여 속성별 가중치 산정에서 의사 결정자의 주관을 배제하고, 적합집단과의 비유사성 산출에서 기여도가 큰 속성만을 선별하여 적용함으로써 속성별 가중치에 합리적인 의미를 부여한다.

추후 연구과제로는 분류된 각 클러스터에서 각각 동일한 값의 속성이 존재하는 경우에 발생할 수 있는 자명해에 대한 처리 방안과 더불어 k-평균 클러스터링과 가중치의 상호 교호작용의 효과에 대한 추가적인 연구가 필요하다.

참 고 문 헌

- [1] Anderberg, R., Cluster analysis for applications, Academic Press, 1973.
- [2] Baumgarten, B., "Optimization of United States Marine Corps Officer Career Path Selection," Master Thesis, Naval Postgraduate School, Monterey CA, 2000.
- [3] Baybars, I., "An Efficient Heuristic Method for the Simple Assembly Line Balancing Problem," International Journal of Production Research, Vol.24(1), pp.149-166, 1986.
- [4] Berrada, M. and E. Stecke, "A Branch and Bound Approach for Machine Load Balancing in Flexible Manufacturing System," Management Science, Vol.32(10), pp.1316-1341, 1986.
- [5] Bobrowski, M. and S. Park, "An Evaluation of Labor Assignment Rules When Workers Are not Perfectly Interchangeable," Journal of Operations Management, Vol.11(7), pp.257-268, 1993.
- [6] Burns, N. and J. Koop, "A Modular Approach to Optimal Multiple-Shift Manpower Scheduling," Operations Research, Vol.35, pp.100-110, 1987.
- [7] Constantopoulos, P., "Decision Support for Massive Personnel Assignment," Decision Support System, Vol.5, pp.355-363, 1989.
- [8] Fayyad, M. et al., Advances in Knowledge Discovery and Data Mining, MIT Press, 1996.
- [9] Grinold, C., "Manpower Planning with Uncertain Requirement," Operation Research, Vol.24(3), pp.177-189, 1976.
- [10] Huang, Z., "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values," Data Mining and Knowledge Discovery, Vol.2, pp.283-304, 1998.
- [11] Klingman, D. and V. Phillips, "Topological and Computational Aspects of Preemptive Multicriteria Military Personnel Assignment Problem," Management Science, Vol.30(11), pp.1362-1375, 1984.
- [12] Liang, T. and J. Lee, "A Systems Approach to Integrate Manpower Planning and Operation," Socio-Economic Planning Sciences, Vol.19(6), pp.371-377, 1984.
- [13] Liang, T. and J. Thompson, "A Large-Scale Personnel Assignment Model for the Navy," Decision Science, Vol.18(2), pp.234-249, 1987.
- [14] Liang, T. and B. Buclatin, "Improving the Utilization of Training Resources through Optimal Personnel Assignment in the U.S. Navy," European Journal of Operational Research, Vol.33, pp.183-190, 1988.
- [15] MacQueen, J., "Some Methods for Classification and Analysis of Multivariate Observations", in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Eds., L. M. Le Cam & J. Neyman, Berkeley: University of California Press, Vol.1, pp.281-297, 1967.
- [16] Makarenkov, V. and P. Legendre, "Optimal Variable Weighting for Ultrametric and

- Additive Trees and K-means Partitioning: Method and Software," *Journal of Classification*, Vol.18, pp.245-271, 2001.
- [17] Naracimhan, R., "An Algorithm for Single Shift Scheduling of Hierarchical Workforce," *European Journal of Operational Research*, Vol.96(1), pp.113-121, 1996.
- [18] Rachamadugu, V. and B. Talbot, "Improving the Equality of Workload Assignments in Assembly lines," *International Journal of Productions Research*, Vol. 29(3), pp.619-633, 1991.
- [19] Sarker, R., "Optimum Manpower Models for a Production System with Varying Production Rates," *European Journal of Operational Research*, Vol.24, pp.447-454, 1986.
- [20] Sweeny, B., "A Network Model(OSGM-NPS) for the U.S. Marine Corps Officer Staffing Goal Problem," Masters Thesis, Naval Postgraduate School, Monterey CA, 1993.
- [21] Sweetser, A., K. Olson, M. Grabski, K. Miller, A. Seise and S. Sanborn, "An Analysis of Personnel Distribution Options for the Chief of Staff, Army," *Phalanx*, Vol.34(3), pp.12-15, 2001.
- [22] Verbeek, J., "Decision Support System - An Application in Strategic Manpower Planning of Airline Pilots," *European Journal of Operational Research*, Vol.55, pp.368-381, 1991.
- [23] Werners, B., "Interactive Multiple Objective Programming Subject to Flexible Constraints," *European Journal of Operational Research*, pp.342-349, 1987.
- [24] Zuidweg, K. and J. van Dam, "On the Optimal Allocation of Crewmen Aboard a Warship," *Naval Research Logistics Quarterly*, Vol.28(3), pp.503-517, 1981.
- [25] 정지원, 최인찬, "A Genetic Algorithm for Clustering in Data Mining," *대한산업공학회/한국경영과학회 2000 춘계공동학술대회 논문집*, pp.306-309, 2000.
- [26] Legendre, P., <http://www.fas.umontreal.ca/biol/legedre/>