

기계학습을 이용한 수출 컨테이너의 무게그룹 분류

강재호
부산대학교 컴퓨터공학과 Post-Doc.
(hkang@pusan.ac.kr)

강병호
부산대학교 컴퓨터공학과 박사과정
(bhokang@pusan.ac.kr)

류광렬
부산대학교 컴퓨터공학과 교수
(krnyu@pusan.ac.kr)

김갑환
부산대학교 산업공학과 교수
(kapkim@pusan.ac.kr)

컨테이너 터미널에서는 장치장(yard)으로 반입되는 수출 컨테이너의 무게를 몇 단계 그룹으로 나누고 각 무게그룹 별로 모아서 장치한다. 이는 수출 컨테이너를 선박에 싣는 적하작업 시 선박의 안정성을 위하여 무거운 무게그룹의 컨테이너들을 장치장에서 먼저 반출하여 선박의 바닥 쪽에 놓기 위함이다. 하지만 반입되는 컨테이너의 무게그룹을 결정할 때 사용하는 운송사로부터 받은 무게정보는 부정확한 경우가 많아 하나의 스택(stack)에 서로 다른 무게그룹에 속하는 컨테이너들이 섞여서 쌓이게 된다. 이로 인하여 무거운 무게그룹의 컨테이너를 반출할 때 해당 컨테이너의 상단에 놓여진 보다 가벼운 무게그룹의 컨테이너들을 임시로 옮겨야 하는 재취급(rehandling, reshuffling)이 발생하게 된다. 적하작업 시 장치장에서 재취급이 빈번히 발생하면 작업이 지연되므로 터미널 생산성 향상을 위해서는 재취급 발생을 가급적 줄여야 한다. 본 논문에서는 기계학습 기법을 적용하여 반입 컨테이너의 무게그룹을 보다 정확히 추정하는 방안을 제안한다. 또한 탐색을 통하여 분류기 생성에 관여하는 비용행렬(cost matrix)을 조정함으로써 재취급 발생을 줄일 수 있는 분류기(classifier)를 생성하는 방안을 함께 소개한다. 실험 결과 본 논문에서 제안하는 방안 적용 시 재취급 발생을 5~7% 정도 줄일 수 있음을 예상할 수 있었다.

논문접수일 : 2005년 5월 게재 확정일 : 2005년 6월 교신저자 : 강재호

1. 서론

컨테이너 터미널의 생산성은 선석에서 컨테이너를 선박에 싣고 내리는 본선작업을 얼마나 효율적으로 수행하느냐에 큰 영향을 받는다. 본선작업은 크게 선박에 실린 수입 컨테이너를 내리는 양하작업과 선박에 수출 컨테이너를 싣는 적하작업 두 가지로 이루어진다. 이 중에서 적하작업을 수행할 때에는 선박의 안정성을 고려하여 무거운 컨테

이너들을 선박의 바닥 쪽에 배치하는 것을 기본 원칙으로 한다. 적하계획을 수립하는 전문가는 이러한 기본 원칙과 대상 컨테이너들이 장치장 내 어디에 놓여져 있는지를 함께 고려하여 작업을 효율적으로 수행할 수 있도록 적하순서를 계획한다. 적하작업은 선박의 바닥 쪽부터 컨테이너를 쌓는 작업이므로 동일 선박 베이(bay)에 실릴 컨테이너들은 장치장에서 연이어 반출된다. 따라서 장치장에서 무거운 순서로 컨테이너를 반출하는 것

* 이 논문은 교육인적자원부 지방연구중심대학육성사업(차세대물류IT기술연구사업단)의 지원에 의하여 연구되었습니다.

이 용이할수록 적하게획 수립이 수월해지고 작업 흐름 또한 원활해진다. 만일 적하순서상 장치장에서 지금 반출하여야 하는 컨테이너의 상단에 나중에 반출할 컨테이너들이 장치되어 있다면, 부득이하게 위에 놓여 있는 컨테이너들을 임시로 옮겨야 하는데 이러한 부가작업을 재취급이라 한다. 재취급이 빈번히 발생하게 되면 적하작업이 지연된다¹⁾. 따라서, 장치장에서 발생하는 재취급을 줄이는 것은 선석에서의 적하작업을 원활히 하여 터미널의 생산성을 향상시키는데 있어 매우 중요하다.

여러 국내 컨테이너 터미널에서는 적하작업 시 장치장에서 발생하는 재취급을 줄이고, 장치장 공간할당 계획과 운영 및 적하작업 계획과 실행의 효율을 높이기 위하여, 수출 컨테이너들을 그 무게에 따라 몇 단계의 그룹으로 나누어 같은 무게그룹에 속하는 컨테이너들을 가능한 모아서 장치한다. 개별 컨테이너의 무게그룹은 해당 컨테이너가 터미널에 반입되기 전에 컨테이너를 운반하는 운송사로부터 전달 받은 전자문서에 포함된 무게 정보(이하 운송사 무게라 칭함)를 이용하여 파악되며, 실제 컨테이너가 반입되면 이 무게그룹을 기준으로 장치할 위치를 결정한다. 하지만, 운송사 무게 정보는 터미널에서 적하게획을 수립할 때 사용되는 선사로부터 받은 정확한 무게 정보(이하 선사 무게라 칭함)와 틀린 경우가 많아 결과적으로 서로 다른 여러 무게그룹에 속하는 컨테이너들이 섞여 쌓이게 되고, 이로 인하여 적하작업 시 상당한 횟수의 재취급이 발생한다.

본 논문에서는 운송사 무게 정보의 부정확성이

적하작업 과정에서 재취급을 얼마나 발생시키는지 추정하는 방안을 소개하고, 무게 이외에 운송사가 함께 제공한 컨테이너 관련 정보를 활용하여 기계학습의 분류 기술을 적용함으로써 무게그룹 분류의 정확도를 향상시킬 수 있는 방안을 제안한다. 또한 무게그룹 분류의 정확도 향상이 재취급 감소로 이어질 수 있도록 분류기 생성에 관여하는 비용행렬을 탐색 기법으로 조정하는 방안을 함께 소개한다.

본 논문은 먼저 2장에서 관련 연구들을 살펴보고, 3장에서는 운송사 무게 정보의 부정확성이 재취급을 얼마나 발생시키는지 추정하는 방안을 소개한다. 4장에서는 재취급 발생을 줄이기 위하여 기계학습 기법을 적용하는 방안을 제시하고, 5장에서는 수집한 데이터를 이용하여 실험한 결과를 정리하여 분석한다. 마지막 6장에서 결론과 향후 연구 과제로 매듭을 짓는다.

2. 관련 연구

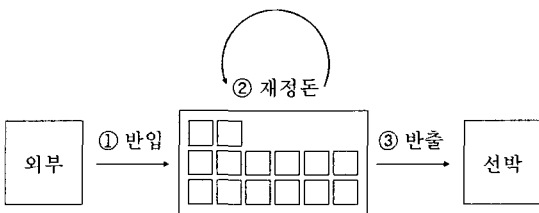
본 장에서는 먼저 장치장 관점에서 컨테이너 이동에 따른 작업 흐름을 소개하고, 관련된 기존 연구들을 요약하여 설명한다.

[그림 1]은 선박에 적하할 수출 컨테이너의 이동 흐름을 간략하게 도식화하여 보여 주고 있다. 수출 컨테이너들은 선박이 입항하기 일정 시점 이전까지 운송사에 의하여 외부에서 터미널로 반입되어 사전에 계획된 장치장 공간에 장치된다²⁾. 일반적으로 장치장 공간은 동일한 호선, 동일한 목적항, 동일한 규격을 가지는 같은 무게그룹의 컨테이

1) (류영옥, 1998)의 연구에 따르면, 4단 6열 베이 구조를 가진 장치장에서 재취급 없이 하나의 컨테이너를 반출하는데 평균 87.9초의 시간이 소요되었으며, 한번의 재취급을 수행하는 데는 평균 74.2초의 시간이 필요하였다.

2) 환적화물과 같은 경우에는 다른 선박의 양하작업에 의하여 장치장으로 반입될 수 있다.

너들을 가능한 모을 수 있도록 계획된다. 이는 목적항과 규격이 동일한 컨테이너들이 선박에서도 같은 베이에 연이어 장치될 가능성이 높으므로 이들 컨테이너들을 한 곳으로 모으는 것이 적하작업의 효율을 높이는데 유리하기 때문이다. 또한 이러한 방식은 장치장 상황을 파악하기가 용이하여 장치장 공간계획과 적하계획을 수립하고 실행하는데 있어서도 효과적이다. 적하계획을 수립하는 전문가가 컨테이너 반입이 완료되면 장치장에서 대상 수출 컨테이너들의 장치위치, 선박 내 공간계획(load profile), 선박의 안정성, 작업할 안벽 크레인(quay crane; QC)의 수 등을 고려하여 적하순서 즉, 장치장에서 컨테이너를 반출하는 순서를 결정한다. 이 때, 동일한 목적항으로 향하는 같은 규격의 컨테이너라면 무게그룹이 적하순서 결정에 가장 큰 영향을 미친다. 해당 선박이 입항하여 적하작업이 시작되면 장치장은 사전에 계획된 순서대로 컨테이너를 반출한다.



[그림 1] 수출 컨테이너의 흐름

적하 작업을 수행하고 있는 ③의 적하를 위한 반출 단계에서 장치장 크레인이 작업 중에 자주 이동하고 재취급을 빈번히 처리해야 하면 작업효율이 떨어지게 된다. 하지만 이 단계에서는 반출을 아무리 효율적으로 처리한다고 하더라도, 계획에 맞춰 적하작업을 수행하고 있는 중이기 때문에 작업 수행시간을 단축하는 데는 한계가 있다. 따라서 적하를 위한 반출 이전 단계인 ① 반입 단계와 ②

장치장에서의 유휴 시간을 활용하여 컨테이너들을 재정돈을 하는 단계에서 이 후의 적하작업을 보다 효율적으로 수행할 수 있게 하는 효과적인 방안이 요구되는 것이다.

①의 반입 단계에서는 적하 시 같은 선박 베이에 실릴 가능성이 높은 컨테이너들을 가능한 동일한 장치장 베이 또는 인접한 장치장 베이에 위치시키는 동시에, 적하순서를 결정하는데 중요하게 고려되는 컨테이너 무게 정보를 활용하여 반입되는 컨테이너의 적절한 장치위치를 결정하는 방식으로 접근한다. 특정 선박에 실릴 컨테이너의 수와 각 목적항으로 향하는 컨테이너의 비율 등을 기존 작업자료로부터 추정하여 사전에 거시적인 장치장 공간계획을 수립한다. 효과적인 장치장 공간계획은 적하작업 시 장치장 크레인의 불필요한 이동을 줄여준다.

적하작업은 선박의 안정성을 위하여 무거운 컨테이너들을 선박의 바닥에서부터 쌓게 되므로 장치장에서는 무거운 컨테이너들을 우선적으로 반출한다. 따라서 반입되는 컨테이너의 무게를 고려하여 장치장 내 장치위치를 결정하면 재취급 발생을 상당히 줄일 수 있다. 특히, 이렇게 장치위치를 효과적으로 결정하여 재취급을 줄이는 방안들은 추가 비용이 거의 들지 않기 때문에 매우 효율적이라 할 수 있다. 반입 컨테이너의 장치위치를 결정하는 방안에 관한 주요 기존 연구들은 다음과 같다.

(Kim et al. 2000)은 재취급 발생을 최소화하기 위하여 수출 컨테이너 반입 시 무게그룹 정보를 이용하여 최적의 장치위치를 결정하는 연구를 수행하였다. 이 연구에서는 서로 다른 무게그룹에 속하는 컨테이너들을 하나의 스택에 쌓는 흔적을 허용하는 상황에서 가벼움, 보통 그리고 무거움의 세 단계로 구분되는 반입 컨테이너의 무게그룹 정보

를 이용하여 동적 계획법(dynamic program)으로 재취급이 최소화되는 장치위치를 선정하는 방안을 제안하였다. 결과의 이해와 적용을 보다 용이하게 하기 위하여 생성된 최적 장치위치 결정 방안을 의사 결정 나무(decision tree) 구조 형태로 표현하는 방법을 함께 제시하였다.

(양지현, 2003)은 제품 창고에서 발생하는 재취급 문제에 대하여 연구하였다. 이 연구에서는 저장 공간의 제약 때문에 혼적을 해야 할 때, 적절한 장치위치 선정의 근거로서 최소 기대 재취급 횟수를 이용하였다. 저장되는 모든 저장 수요들의 입출고 시간, 입출고 순서, 제품 수 등의 사전 정보를 알고 있는 정적인 의사 결정 문제와 사전 정보가 주어지지 않는 동적인 의사 결정 문제로 나누어 접근하였다. 동적 의사 결정 문제를 해결하기 위하여 재취급에 영향을 주는 요소들을 고려한 휴리스틱을 제시하였다. (강재호 et al. 2004b)는 이 연구에서 제안된 휴리스틱을 기반으로 컨테이너 터미널 장치장에서 반입 컨테이너의 무게 정보를 활용하여 장치위치를 결정하는 방안을 제시하였다.

반입 시 무게나 반출 순서 정보를 활용하여 장치위치를 결정하는 기존 연구들은 대부분 결정에 사용되는 정보가 정확하다는 가정하에 출발하였다. 이에 비해 본 논문에서는 현실적으로 장치위치 결정 시 사용되는 운송사 무게 정보가 부정확할 수 있다는 점을 고려하여 접근하고 있다.

[그림 1]의 ②에 해당되는 장치장에서의 유휴 시간을 활용한 재취급 최소화 연구로는 (Kim et al. 1998, 강재호 et al. 2004a, 강재호 et al. 2005, 오명섭 et al. 2005)이 있다. 이들 연구에서는 적하를 고려하여 유휴 시간에 컨테이너들을 효율적으로 재재동 할 수 있는 계획을 수립하는 방안을 제시하였다.

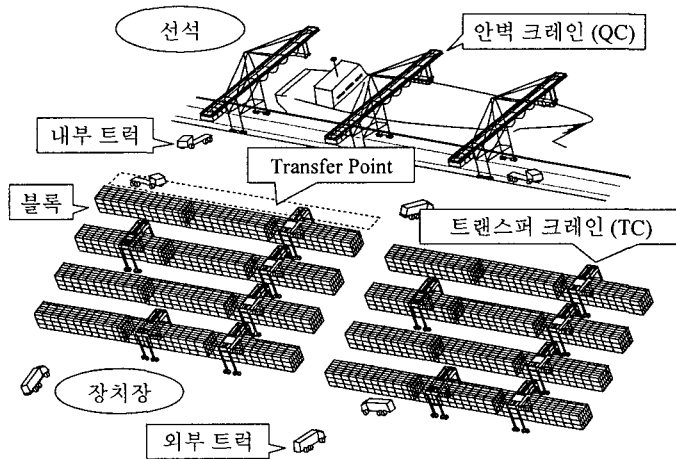
3. 재취급 발생 횟수 추정 방안

본 장에서는 컨테이너 터미널에 대한 간략한 소개와 함께 수출 컨테이너에 대한 운송사 무게 정보가 부정확함으로써 발생하는 재취급 횟수를 추정하는 방안에 대하여 자세히 설명한다.

3.1 컨테이너 터미널

컨테이너 터미널에서 본선작업과 관련한 부분의 구조를 살펴보면 대략 [그림 2]와 같다. 터미널은 크게 해측의 선석과 육측의 장치장으로 구성된다. 해측에는 하나 이상의 선박이 동시에 정박할 수 있는 안벽이 있다. [그림 2]는 하나의 선박이 정박한 경우를 도시하고 있다. 안벽에는 선박으로부터 수입 컨테이너를 내리거나 선박에 수출 컨테이너를 싣는 QC들이 있다. QC는 내부 트럭과 선박 사이에서 컨테이너를 옮긴다. 장치장은 수출입 컨테이너를 적재해두는 곳으로 트랜스퍼 크레인(transfer crane; TC)을 이용하여 내부 트럭에 컨테이너를 싣거나 내린다. 장치장과 QC 사이의 컨테이너 운반은 내부 트럭이 담당한다. 장치장은 여러 개의 블록으로 구성되며 각 블록마다 트럭들이 정차할 수 있는 공간으로 TP(transfer point)가 있다. 내부 트럭이 TP에 도착하면 TC는 내부 트럭에 수출 컨테이너를 싣거나 양하된 수입 컨테이너를 내부 트럭에서 집어 블록 내에 장치하게 된다.

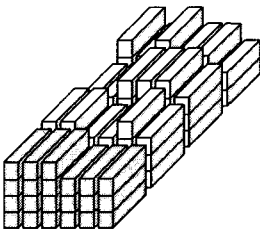
일반적으로 컨테이너 터미널의 생산성은 QC 작업이 얼마나 원활히 이루어지는가에 달려 있다. 즉, QC가 지연 없이 작업할 때 생산성이 높아지게 된다. 따라서 QC에서 작업 지연이 발생하지 않도록 장치장에서 수출입 컨테이너를 처리할 수 있어야 한다. 장치장 관점에서 양하작업은 미리 할당된 공간에 컨테이너를 쌓는 작업이므로 특별한 지연



[그림 2] 컨테이너 터미널

없이 처리가 가능하나, 적하작업은 컨테이너들의 장치위치에 따라 재취급이 발생할 수 있으며 이로 인하여 작업효율이 떨어질 수 있다.

장치장에는 [그림 3]과 같은 블록(block)이 여럿 있으며, 각 블록은 TC의 본체가 움직이지 않고 컨테이너를 장치할 수 있는 공간인 베이들로 구성되어 있다. 각 베이는 여러 개의 스택으로 이루어지는데, 하나의 스택은 장치 높이만 다른 컨테이너 저장 공간의 집합이다. 베이 내에서 스택의 위치를 열(row)이라 하며, 스택에서 컨테이너의 장치위치를 단(tier)이라 한다. [그림 3]은 컨테이너가 장치된 4단 6열 구조의 블록 일부분을 보여주고 있다. 그림에서 음영으로 표시된 컨테이너들은 동일한

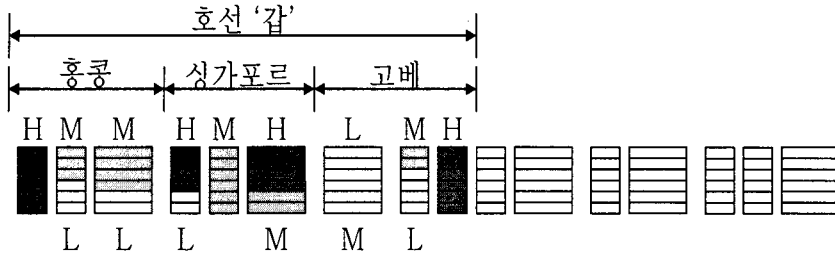


[그림 3] 컨테이너가 장치된 블록의 일부

베이에 속한다. 다음절에서는 적하작업 시 운송사 무게 정보가 부정확하여 발생하는 재취급에 대하여 보다 자세히 소개한다.

3.2 운송사 무게 정보의 부정확성

터미널들은 동일한 호선, 목적항, 규격 그리고 무게그룹을 가진 컨테이너들이 모이도록 장치장 공간을 계획하며, [그림 4]에 이러한 장치장 공간 계획의 예를 보이고 있다. 이 그림은 하나의 블록을 위에서 본 그림이다. 짧은 사각형은 20피트 컨테이너이며 긴 사각형은 40피트 컨테이너이다. 그림에서 홍콩, 싱가포르, 고베는 목적항이며, L(Light), M(Medium), H(Heavy)는 각각 가벼움, 중간, 무거움에 해당되는 컨테이너 무게그룹을 의미한다. 무게그룹은 컨테이너가 터미널에 반입되기 전에 운송사에서 터미널로 보내진 운송사 무게 정보를 각 터미널마다 기준을 가지고 적절히 나누는 것이다. 여기서는 무게그룹을 3단계로 나누는 예를 보이고 있지만 더 세분화될 수도 있다.

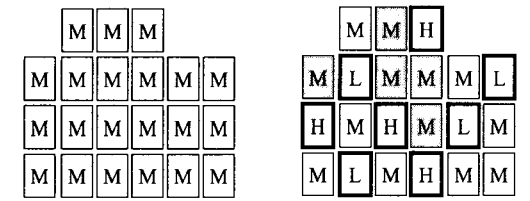


[그림 4] 장치장 계획의 예

운송사 무게 정보가 언제나 정확하다면 적하작업 시 컨테이너 재취급은 발생하지 않겠지만³⁾, 현실적으로 상당수의 운송사 무게 정보에는 오류가 존재한다. [그림 5]에 이러한 예를 보이고 있다. [그림 5]의 (a)는 운송사 무게 정보를 근거로 무게 그룹 M으로 분류된 컨테이너들만 모아서 장치한 4단 6열 구조의 베이 하나를 보이고 있다. 모든 컨테이너들이 반입된 이후에 선사로부터 건네 받은 선사 무게 정보로 해당 베이를 확인해 보면 (b)와 같이 일부 컨테이너들의 무게그룹이 잘못된 것을 확인할 수 있게 된다. 결과적으로 이 베이에는 본래 의도하였던 무게그룹 M 이외의 다른 무게그룹

들에 속하는 컨테이너들이 상당수 섞이게 된다. 그림에서 이러한 컨테이너들은 굵은 사각형으로 표시하였다. 일부 컨테이너들은 자신들보다 무거운 그룹에 속하는 컨테이너가 하단에 있어 적하작업 시 재취급 되어야 한다. 그림에서 반드시 재취급 되어야 하는 컨테이너들은 회색으로 표시하였다. 운송사 무게 정보가 부정확한 이상 이러한 재취급 발생을 피할 수 없다.

운송사 무게 정보와 선사 무게 정보의 불일치 정도를 보다 자세히 살펴보기 위하여 신선대 컨테이너 터미널⁴⁾에서 2004년도 6월부터 10월 사이에 처리한 20피트 수출 컨테이너들 중 1,000개를 선택하여 [그림 6]과 같이 그래프에 표시하였다⁵⁾. 여기서 그림을 간결하게 나타내기 위하여 운송사 무게 또는 선사 무게가 30톤 이하인 컨테이너들만 표시하였다. 그래프에서 가로축은 운송사 무게이며 세로축은 선사 무게이다. 그래프의 각 점은 하나의 컨테이너에 해당된다. 점선들은 컨테이너 무게그룹을 나누는 기준으로 각 축에서 무게 0톤에 가까운 첫 번째 점선이 무게그룹 L과 M을 나누고,



(a) 운송사 무게 정보로 장치한 상태의 예 (b) 선사 무게 정보로 확인된 상태의 예

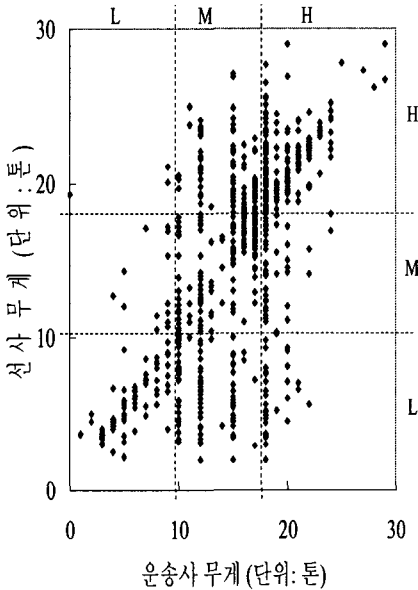
[그림 5] 선사 무게 정보로 확인하기 전과 확인 후의 베이 상태의 예

3) 장치장 공간 부족으로 인해 다른 무게그룹의 컨테이너들이 있는 스택에 반입된 컨테이너를 장치하여야 하는 경우나 이미 장치된 컨테이너의 호선이나 목적항이 변경되는 경우 재취급이 발생할 수 있다.

4) <http://www.shinsundae.co.kr/>

5) 냉동, 위험물 등과 같은 특수 컨테이너들과 자(自)부두에서 환적하여 수출하는 컨테이너들은 전용 장치장 공간이 할당되어 터미널 외부에서 반입되는 일반 수출 컨테이너들과 그 처리 과정이 일부 다르므로 분석과 이후 실험 대상에서 제외하였다.

두 번째 점선이 무게그룹 M과 H를 구분한다. 그림에서와 같이 운송사 무게 정보와 선사 무게 정보 간에 상당한 불일치가 있음을 확인할 수 있다.



[그림 6] 20피트 컨테이너의 무게 분포

<표 1>은 운송사 무게그룹(이하 운송사 무게 그룹)과 선사 무게 정보를 이용하여 분류한 무게 그룹(이하 선사 무게그룹)과의 일치 정도를 나타내고 있다. 이 표에서 운송사 무게 정보를 신뢰하여 무게그룹을 분류하면 정확도가 70%를 넘지 못함을 알 수 있다. 20피트 컨테이너에 비해 40피트 컨테이너에 대한 무게그룹 분류의 정확도가 더 낮았다.

<표 1> 운송사 무게로 분류한 무게그룹의 정확도

컨테이너 크기	무게그룹의 정확도
20피트	69.3%
40피트	64.8%

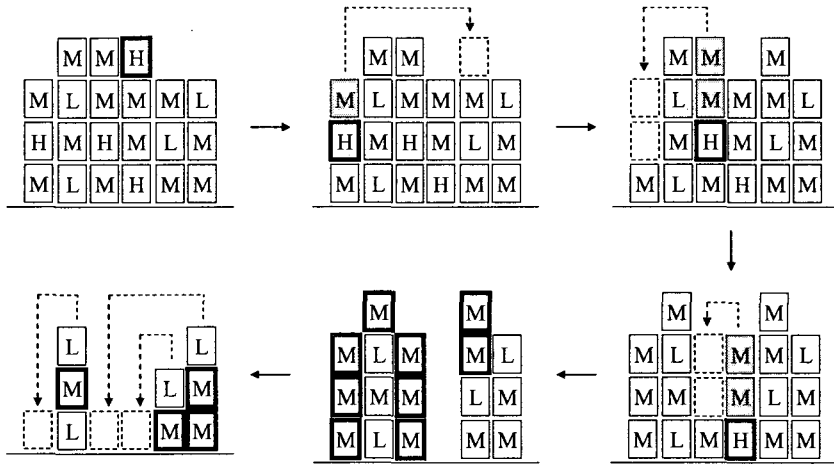
<표 2>와 <표 3>은 각각 20피트와 40피트 컨테이너들을 대상으로 운송사 무게그룹과 선사 무게그룹 간에 쌍으로 이루어진 모든 가능한 조합에 대하여 컨테이너 발생 빈도를 표현한 혼돈행렬(confusion matrix)이다. 이러한 혼돈행렬은 [그림 6]과 같이 표현된 컨테이너 무게 분포 그래프로부터 무게그룹 분류 기준, 즉 점선으로 구분된 각 영역 별 컨테이너의 비율을 구하면 얻을 수 있다. 예를 들어 <표 2>에서 회색으로 표시된 7.1%는 20피트 컨테이너들 중에는 운송사 무게그룹으로는 M으로 분류되어 장치장에 장치되었지만, 선사 무게그룹으로는 L에 해당되는 컨테이너가 7.1%에 달한다는 것을 의미한다. 표에서 운송사와 선사간의 무게그룹 비율도 상당한 차이가 있음을 확인할 수 있다. <표 1>에서 40피트 컨테이너의 경우 20피트 컨테이너에 비해 분류 정확도가 더 낮았는데, <표 2>와 <표 3>을 비교하여 살펴보면 40피트 컨테이너의 경우 분류하기 까다로운 무게그룹 M에 해당되는 컨테이너의 비율이 상대적으로 높은 것을 확인할 수 있다.

<표 2> 20피트 컨테이너의 혼돈행렬

운송사 무게그룹	선사 무게그룹		
	L (20.0%)	M (23.1%)	H (56.9%)
L (13.9%)	11.1%	1.9%	0.9%
M (39.7%)	7.1%	17.4%	15.2%
H (46.4%)	1.8%	3.8%	40.8%

<표 3> 40피트 컨테이너의 혼돈행렬

운송사 무게그룹	선사 무게그룹		
	L (24.2%)	M (46.2%)	H (29.6%)
L (13.8%)	9.6%	3.8%	0.5%
M (59.9%)	12.8%	36.6%	10.5%
H (26.2%)	1.8%	5.8%	18.6%



[그림 7] 적하 시 컨테이너 재취급을 고려한 반출의 예

3.3 재취급 횟수 추정 방안

이 절에서는 컨테이너 무게 정보의 부정확성으로 인하여 발생하는 재취급에 대하여 설명하고, 혼돈행렬 정보를 이용하여 재취급 횟수를 추정하는 방안을 소개한다.

[그림 5]의 (b)와 같이 장치된 상황에서 적하작업이 진행되어 컨테이너들이 반출되는 순차적인 과정을 [그림 7]에 보이고 있다. 무거운 컨테이너들을 먼저 적하한다고 가정한다면, 무게그룹이 H, M, L인 순서로 컨테이너들을 장치장에서 반출하게 된다⁶⁾. 이 과정에서 하단에 자신보다 일찍 반출되어야 하는 컨테이너가 있는 컨테이너는 반드시 재취급 되어야 한다. [그림 5]의 (b)를 살펴보면 이러한 경우에 해당되는 컨테이너는 모두 8개가 있으므로 최소 8번의 재취급이 발생한다. 반드시 재취급 되어야 하는 컨테이너만을 세었기 때문에

이렇게 구한 재취급 횟수는 하한값(lower bound)이 된다. [그림 7]은 하한값에 해당되는 횟수의 재취급으로 베이 내 모든 컨테이너를 반출하는 예를 보이고 있다. 각각의 상태에서 반출하고자 하는 컨테이너들은 굵은 테두리로 표현하였고, 재취급 되어야 하는 컨테이너들은 회색 사각형으로 나타내었다. 대개의 경우 하한값만큼의 재취급으로 한 베이 내의 컨테이너들을 모두 반출할 수 있으므로, 본 논문에서는 문제 모형을 단순화하기 위하여 재취급 된 컨테이너는 다시 재취급 되지 않는다고 가정하였다.

이어서 혼돈행렬 정보를 이용하여 컨테이너 재취급 횟수를 추정하는 방안에 대한 간단한 예를 보이고자 한다. 먼저 동일한 운송사 무게그룹으로 분류된 컨테이너들 중에서 선사 무게그룹 별 상대적 비율을 분석해 보면 <표 4>와 같은 행렬을 얻을 수 있다. 예를 들어 표에서 회색으로 표시된 20%는 운송사 무게그룹 L로 분류된 컨테이너들 중에서 20%가 선사 무게그룹으로는 H에 해당하였다는 의미이다. 여기서는 예제의 계산 편의를 위하여 확률을 임의로 설정하였다. 이러한 표는 앞에

6) 적하계획을 수립하거나 수행하는 과정에서 작업의 효율을 위하여 안전한 범위 내에서 무게그룹에 의한 적하순서상의 제약을 일부 어길 수 있으나, 본 논문에서는 이를 고려하지 않는다.

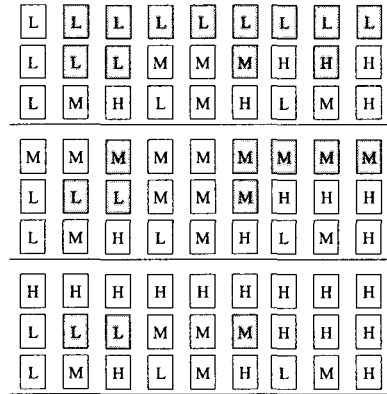
서 소개한 혼돈행렬 정보를 간단하게 가공하여 얻을 수 있다.

<표 4> 운송사 무게그룹 별로 정규화한 혼돈행렬의 예

운송사 무게그룹	선사 무게그룹		
	L (33.3%)	M (33.3%)	H (33.3%)
L (33.3%)	60%	20%	20%
M (33.3%)	30%	50%	20%
H (33.3%)	10%	30%	60%

하나의 스택에 3단으로 컨테이너를 쌓는 방법의 수는 [그림 8]과 같이 총 27가지가 될 것이다. 그림에서 회색으로 표시된 컨테이너들은 반출 시 재취급 되어야 한다. 운송사 무게그룹으로 장치된 스택이 하나 주어지면 그 스택이 실제 선사 무게그룹 기준으로 이들 27가지 중 어떤 것에 해당할지 그 각 경우에 대한 확률을 구할 수 있다. 예를 들어 하나의 스택에 운송사 무게그룹으로 하단부터 LLM으로 쌓았다고 가정하자. 이 스택이 선사 무게그룹으로는 MLH 상태일 가능성은 다음과 같이 계산된다. 첫 번째 단의 컨테이너가 선사 무게그룹으로 M일 가능성 0.2, 두 번째 단의 컨테이너가 선사 무게그룹 L일 가능성 0.6, 세 번째 단의 컨테이너가 선사 무게그룹 H일 가능성 0.2를 곱하면 MLH 장치 상태 발생 확률은 0.024가 된다. 예로 든 계산과 관련된 확률은 표 4에 굵은 글씨로 나타내었다. 이 상태에서는 두 번째 단의 컨테이너(선사 무게그룹 L)는 하단에 자신보다 무거운 무게그룹의 컨테이너(선사 무게그룹 M)가 존재하므로 반출 시 1회의 재취급이 발생하게 된다. 이러한 방식으로 모든 발생 가능한 경우를 대상으로 발생 확률과 재취급 횟수를 구하고, 구해진 재취

급 횟수를 각각의 발생 확률로 가중합하면 운송사 무게그룹 LLM인 스택의 예상 재취급 횟수를 구할 수 있다.



[그림 8] 3단 장치 시 발생 가능한 27가지 장치 상태

<표 5>는 <표 2>와 <표 3>의 혼돈행렬을 이용하여 동일한 운송사 무게그룹으로 분류된 컨테이너들만 하나의 스택에 쌓는 즉, 혼적을 허용하지 않고 장치할 때, 하나의 컨테이너가 재취급 될 확률을 구한 것이다. 예를 들어 컨테이너를 3단으로 쌓는다면 혼적을 허용하지 않으므로 운송사 무게그룹으로는 LLL, MMM, HHH의 세 가지 경우가 나타날 수 있다. 각각의 경우에 대하여 앞에서 소개한 방식으로 예상되는 재취급 수를 구한 후, 혼돈행렬에 표현되어 있는 운송사 무게그룹 L, M, H에 해당되는 컨테이너의 발생 확률을 곱하여 가중합하면 한 스택에서 발생할 것으로 예상되는 재취급 횟수를 추정할 수 있다. 표 5는 이렇게 구한 값들을 장치 단수로 나눈 것이다. 20피트 컨테이너의 경우 3단으로 쌓는다면 스택 내 하나의 컨테이너는 약 15% 확률로 재취급 된다. 따라서 하나의 스택에서는 약 0.45회 (= 0.15 × 3), 하나의 3단 6열 베이에서는 약 2.7회 (= 0.45 × 6)의 재취급 발생이

예상된다. 장치 단수가 높아질수록 하나의 컨테이너가 재취급 될 확률은 증가함을 확인할 수 있는데, 이는 컨테이너를 고단으로 장치하는 경우 운송사 무게 정보와 장치위치 결정에 더욱 신경을 써야 함을 의미한다. 상대적으로 운송사 무게그룹의 정확도가 떨어지는 40피트 컨테이너의 경우 재취급이 더 빈번히 발생할 것임을 예상할 수 있다.

<표 5> 장치 단수 별 컨테이너의 평균 재취급 확률

컨테이너 크기	개별 컨테이너의 재취급 확률			
	3단	4단	5단	6단
20피트	0.150	0.187	0.215	0.236
40피트	0.207	0.263	0.307	0.343

4. 학습을 이용한 컨테이너 무게그룹 분류

본 장에서는 기계학습의 분류 기술을 이용하여 컨테이너 무게그룹을 보다 정확하게 분류하고, 이를 활용하여 재취급 발생 횟수를 줄이는 방안을 소개한다.

4.1 학습을 이용한 무게그룹 분류

기계학습의 분류 기술을 적용하면 분류하고자 하는 각 대상과 관련된 정보(여기서는 운송사가 전송한 컨테이너 관련 정보)를 바탕으로 몇 개의 카테고리(category 또는 class, 여기서는 무게그룹) 중의 하나로 사상(mapping)할 수 있는 함수를 주어진 예제들을 이용하여 생성할 수 있다(Witten and Frank, 1999). 반입되는 컨테이너의 무게그룹을 분류하는데 사용하는 운송사 무게 정보가 있음에도 추가의 정보를 활용하여 분류 정확도를 향상

시킬 수 있는 이유는 운송사에 따라 제공한 무게 정보의 신뢰 정도와 무게 정보에 대한 편향이 다르기 때문이다⁷⁾. 운송사들은 경우에 따라 컨테이너 무게를 측정하지 않고 경험상 추정되는 무게를 터미널로 전송한다. 따라서 운송사에 따라 동일한 컨테이너라도 서로 다르게 무게를 예상할 가능성이 있다. 예를 들어 운송사 A는 상대적으로 운송사 B에 비하여 컨테이너 무게를 과하게 기입하거나 또는 그 반대의 경우가 있을 수 있다. 따라서 해당 컨테이너를 반입 요청한 운송사가 어디인지, 컨테이너 종류는 어떤 것인지와 같은 정보를 추가로 활용하여 무게그룹을 분류한다면, 단순히 운송사에서 제공한 무게 정보를 신뢰하여 무게그룹을 분류하는 경우에 비해 그 정확도를 높일 수 있다.

본 연구에서 이러한 목적으로 수집한 컨테이너 관련 정보는 <표 6>과 같다. 표에서의 예는 운송사에서 컨테이너 반입 요청 시 21톤으로 무게를 기입하여 터미널에 알려 주었기에 해당 컨테이너를 무게그룹 H로 분류하여 장치하였지만, 선사 무게 정보를 확인한 결과 이보다 무게가 적게 나가 무게그룹 M으로 재분류한 경우이다.

<표 6> 학습에 사용한 컨테이너 관련 정보

항목	설명	예
운송사명	운송사 이름	ABCDE
크기	컨테이너 크기	20피트
종류	컨테이너 종류	일반
환적 여부	타(他)부두 환적 여부	FALSE
운송사 무게	운송사 무게	21 (톤)
선사 무게그룹	선사 무게그룹	M

7) 엄밀하게는 운송사 직원 별로 이러한 신뢰 정도, 경험 축적 및 기준이 다르다고 할 수 있으나, 본 연구에서 수집한 데이터에서는 직원별 구분이 어려워 운송사별로 다르다고 가정하였다.

4.2 비용행렬의 탐색

분류기(classifier)를 생성하는 학습 알고리즘은 일반적으로 분류 시 오류가 최소화되는 분류기를 생성하고자 한다. 하지만, 컨테이너 무게그룹 분류 문제는 단순히 더 정확하게 무게그룹을 분류한다고 하여 반드시 재취급 수를 줄일 수 있는 것은 아니다. 이해를 돕기 위하여 예를 들어 설명하고자 한다.

<표 7>과 <표 8>은 각각 어떤 분류기 h_1 과 h_2 를 이용하여 컨테이너 무게그룹을 분류한 결과를 혼돈행렬로 나타낸 것이다. 선사 무게그룹으로 L, M, H의 비율이 모두 1/3로 동일하다면 이 두 분류기는 모두 70%의 정확도를 가진다. 분류기 h_2 를 분류기 h_1 과 비교한다면 분류기 h_2 는 무게그룹 L과 M은 잘 분류하는데 비해, 무게그룹 H에 대해서는 상대적으로 오류가 많은 편이다. <표 9>는 이 두 분류기가 생성한 혼돈행렬을 이용하여 3장에서 소개한 방안으로 혼적을 허용하지 않을 때 장치 단수 별로 하나의 컨테이너가 재취급 될 확률을 추정한 결과이다. 동일한 단수로 장치 시 더 나은 성능을 보인 경우를 굵은 글씨로 표기하였다. 표에서 알 수 있듯이 컨테이너를 3단 또는 4단까지 장치하는 경우에는 분류기 h_2 를, 5단 또는 6단까지 장치하는 경우에는 분류기 h_1 을 이용하는 것이 예상되는 재취급 횟수를 줄일 수 있음을 알 수 있다.

<표 7> 분류기 h_1 로 생성한 혼돈행렬

h_1 이 예상한 무게그룹	선사 무게그룹		
	L	M	H
L	23.3%	5.0%	5.0%
M	5.0%	23.3%	5.0%
H	5.0%	5.0%	23.3%

<표 8> 분류기 h_2 로 생성한 혼돈행렬

h_2 가 예상한 무게그룹	선사 무게그룹		
	L	M	H
L	26.7%	3.3%	3.3%
M	3.3%	26.7%	13.3%
H	3.3%	3.3%	16.7%

<표 9> 장치 단수 별 컨테이너의 평균 재취급 확률

분류기	개별 컨테이너의 재취급 확률			
	3단	4단	5단	6단
h_1	0.191	0.246	0.289	0.325
h_2	0.186	0.244	0.290	0.328

이 예에서 컨테이너 무게그룹 분류 문제와 관련하여 두 가지 특성을 파악할 수 있다. 첫 번째는 무게그룹 분류의 정확도와 예상 재취급 수가 반드시 반비례하지 않는다는 점이다. 즉 정확도가 더 우수한 분류기를 만들 경우 예상 재취급 수가 줄 가능성이 높겠지만, 반드시 그러하리라 확신할 수는 없다. 두 번째는 컨테이너 장치 단수를 몇 단으로 하느냐에 따라 가장 효과적인 분류기가 다르다는 점이다. <표 9>에서 이러한 점을 확인할 수 있다. 이러한 현상이 발생하는 이유는 컨테이너 무게그룹 분류 문제가 어떤 무게그룹에 속하는 컨테이너를 어느 무게그룹으로 잘못 분류했느냐에 따라 발생하는 재취급 횟수가 다르기 때문이다. 예를 들어 무게그룹 M에 속하는 어떤 컨테이너 e_M 를 무게그룹 H로 잘못 분류하는 경우에는 반출 시 잘못 분류된 e_M 하나만 재취급하면 된다. 이에 비해 e_M 을 무게그룹 L로 잘못 분류하면 반출 시 e_M 상단에 있는 무게그룹 L에 속하는 컨테이너들을 모두 재취급하여야 한다. 따라서 e_M 을 어느 카테고리

잘못 분류하느냐에 따라 그리고 컨테이너를 몇 단으로 쌓느냐에 따라서 재취급 횟수가 달라지게 된다. 기계학습에서는 이러한 특성을 가진 문제를 비용민감(cost-sensitive) 분류 문제라 한다(Domingos, 1999).

비용행렬은 분류기가 실제 카테고리 c_1 인 예제를 다른 카테고리 c_2 로 잘못 분류함으로써 발생하는 비용을 각 카테고리 조합에 대하여 행렬로 표현한 것이다. 이러한 비용행렬을 생성하는 방법을 예를 들어 설명한다. <표 10>은 금융 기관에서 대출을 승인 또는 거절하여야 하는 문제의 이익행렬(profit matrix)을 보이고 있다. 만일 대출자가 대출금을 상환할 능력을 가지고 있을 때(승인) 분류기가 대출을 허가한다면(승인) 금융 기관은 1만원의 이자 수익을 얻을 수 있을 것이며, 반대로 대출금을 상환할 능력이 없음에도(거절) 대출을 허가한다면(승인) 10만원의 손해를 본다고 하자. 문제를 간단히 하기 위하여 대출 심사에는 비용이 들지 않고, 대출을 승인하는 것이 바람직한 고객의 수와 거절하는 것이 바람직한 고객의 수는 일정하다고 가정한다면⁸⁾ <표 10>은 <표 11>과 같이 표현될 수 있다⁹⁾. <표 10>에서

8) 해당 금융기관이 대출을 얼마나 까다롭게 하느냐에 관계없이 대출을 희망하는 고객의 수와 고객들의 카테고리의 비율은 언제나 동일하다는 가정을 의미한다.

9) <표 10>과 <표 11>은 학습 알고리즘이 최적화(여기서는 이익행렬이므로 최대화)하여야 하는 동일한 목적함수를 다르게 표현한 것이다. 분류 결과로 $\langle(n_{11}, n_{12}), (n_{21}, n_{22})\rangle$ 과 같은 혼돈행렬을 얻었다면, <표 10>은 목적함수 $n_{11} - 10n_{12}$ 를 의미하며, <표 11>은 목적함수 $-n_{21} - 10n_{12}$ 를 의미한다. 실제 카테고리 각각에 속하는 고객의 수 $n_{11} + n_{21}$ 과 $n_{12} + n_{22}$ 는 고정이라고 가정하였다. 이를 각각 N_1, N_2 라 두자. 따라서 <표 11>의 목적함수는 $(n_{11} - N_1) - 10n_{12} = n_{11} - 10n_{12} - N_1$ 로 다시 쓸 수 있으므로 두 목적함수는 상수만큼 차이 나게 된다.

는 대출을 승인하여야 할 고객에게 대출을 허가함으로써 1만원의 이익을 발생한다고 표현하였고, <표 11>에서는 대출을 승인하는 것이 바람직한 고객을 거절함으로써 1만원의 손해를 본다고 표현하였다. <표 12>는 <표 11>을 이익 대신 손해의 관점에서 다시 표현한 비용행렬이다. 학습 알고리즘은 이러한 형태로 비용행렬이 주어지면 전체 비용이 최소화되는 분류기를 생성하고자 노력하게 된다.

<표 10> 대출 승인 문제에 대한 이익행렬의 예 1

분류기의 선택	실제 카테고리	
	승인	거절
승인	1만원	-10만원
거절	0만원	0만원

<표 11> 대출 승인 문제에 대한 이익행렬의 예 2

분류기의 선택	실제 카테고리	
	승인	거절
승인	0만원	-10만원
거절	-1만원	0만원

<표 12> 대출 승인 문제에 대한 비용행렬의 예

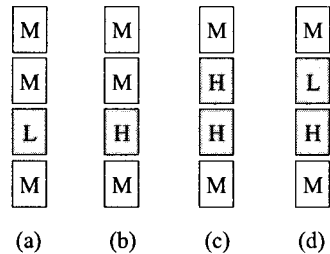
분류기의 선택	실제 카테고리	
	승인	거절
승인	0만원	10만원
거절	1만원	0만원

예로 들었던 대출 승인 문제의 경우 해당 금융 기관의 기존 영업 실적을 분석함으로써 잘못 분류한 경우에 발생하는 비용을 추정하기가 비교적 수

일하지만¹⁰⁾, 컨테이너 무게그룹 분류의 경우 이러한 비용행렬을 계산해내기가 쉽지 않다. 간단한 비용 추정 방법의 하나로 [그림 8]과 같이 발생 가능한 장치 상태 각각에 대하여 재취급이 발생한 요인을 분석하여 비용행렬을 생성할 수 있다. 하지만 이 방법을 적용하고자 할 때 아래와 같이 재취급 발생 요인을 분석하기 까다로운 경우가 발생한다.

[그림 9]는 무게그룹 M으로 분류된 컨테이너들만으로 4단을 쌓았을 때 일부 컨테이너의 무게그룹이 잘못 분류된 몇 가지 경우를 보이고 있다. 그림에서 회색으로 표시된 컨테이너들은 잘못 분류된 컨테이너들이다. [그림 9]의 (a)는 무게그룹 L인 컨테이너를 무게그룹 M으로 잘못 분류하였기에 1회의 재취급이 발생하는 경우이다. 따라서 이 경우 무게그룹 L을 무게그룹 M으로 잘못 분류하면 1의 비용이 발생했다고 할 수 있다. (b)는 무게그룹 H인 컨테이너를 무게그룹 M으로 잘못 분류하여 2회의 재취급이 발생하였으므로 무게그룹 H를 무게그룹 M으로 잘못 분류하면 2의 비용이 발생한다고 할 수 있다. (c)는 무게그룹 H에 속하는 2개의 컨테이너를 무게그룹 M으로 분류하였는데 재취급 발생 횟수는 1회이다. 잘못 분류한 컨테이너의 수가 2개이므로 재취급 횟수를 절반씩 나누어 무게그룹 H를 각각 무게그룹 M으로 잘못 분류한 경우 0.5의 비용이 발생한다고 추정한다면 (b)와 형평성이 맞지 않게 된다. 즉, 잘못 분류한 컨테이너가 더 많음에도 비용이 적게 발생한 것이다. (d)는 서로 다른 두 무게그룹 L과 H에 속하는 컨테이너들을 무게그룹 M으로 잘못 분류한 경우이

다. 이 경우 무게그룹 L에 속하는 컨테이너를 제대로 분류하였다고 하더라도 재취급 횟수는 2회로 동일하므로 무게그룹 L과 무게그룹 H에 해당되는 컨테이너를 잘못 분류한 각각의 경우가 얼마만큼 재취급을 발생시켰는지 판정하기가 쉽지 않다.



[그림 9] 무게그룹 분류 문제에서 비용 설정의 모호함

컨테이너 무게그룹 분류 문제는 생성된 분류기를 검증(validation)하면 해당 분류기의 혼돈행렬을 추정할 수 있으며, 이렇게 얻어진 혼돈행렬을 이용하여 예상 재취급 횟수를 계산할 수 있다¹¹⁾. 따라서 본 논문에서는 비용행렬을 직접 생성하기 보다는 탐색을 통하여 적절한 비용행렬을 발견하는 방식으로 접근하였다. 표준 비용행렬에서 대각요소를 제외한 나머지는 변경이 가능한 변수로 두고 언덕 오르기 탐색(hill-climbing search) 기법을 적용하였다. 비용행렬을 탐색하는 알고리즘은 [그림 10]에 나타내었다.

11) 추정된 혼돈행렬의 각 항목을 확률로 나타내어 $\langle (p_{11}, p_{12}, p_{13}), (p_{21}, p_{22}, p_{23}), (p_{31}, p_{32}, p_{33}) \rangle$ 와 같이 표현하고 혼적을 허용하지 않는 상황에서 컨테이너를 2단으로 쌓는다고 가정하자. $p_i = \sum_j p_{ij}$ 라 두면 재취급 횟수는 $r^* = \sum_i (p_{12} \times p_{11} + p_{13} \times p_{11} + p_{13} \times p_{22}) / p_i$ 로 추정할 수 있다. 이 수식이 학습 알고리즘이 최소화하여야 하는 목적함수이며 앞의 금융기관에서의 대출 승인 문제의 목적함수와는 달리 비선형이기 때문에 각각의 오분류에 대한 비용을 직접적으로 추정하기 어려운 것이다.

10) 엄밀하게 말한다면 은행 대출 승인도 비용행렬이 고정된 문제가 아니다. 대출 허가가 까다로워지면 대출 요청 자체가 줄어들게 될 것이므로 기업 운영에 필요한 고정 비용을 고려하게 되면 잘못 분류한 경우에 대한 비용은 분류기에 따라 변하게 된다.

```

procedure Search_CM (iCM, tSet)
    iCM : initial cost matrix
    tSet : training set

    cCM ← iCM
    cH ← generate_classifier (cCM, tSet)
    cEval ← evaluate_classifier (cH, tSet)

    repeat
        improved ← false
        nCMset ← generate_neighbors (cCM)
        foreach nCM in nCMset
            nH ← generate_classifier (nCM, tSet)
            nEval ← evaluate_classifier (nH, tSet)
            if (nEval < cEval)
                improved ← true
                cCM ← nCM
                cH ← nH
                cEval ← nEval
            break foreach
        end if
    end foreach
    until not improved
    return cCM, cH, cEval
end procedure
    
```

[그림 10] 비용행렬 탐색 알고리즘

5. 실험 결과 및 분석

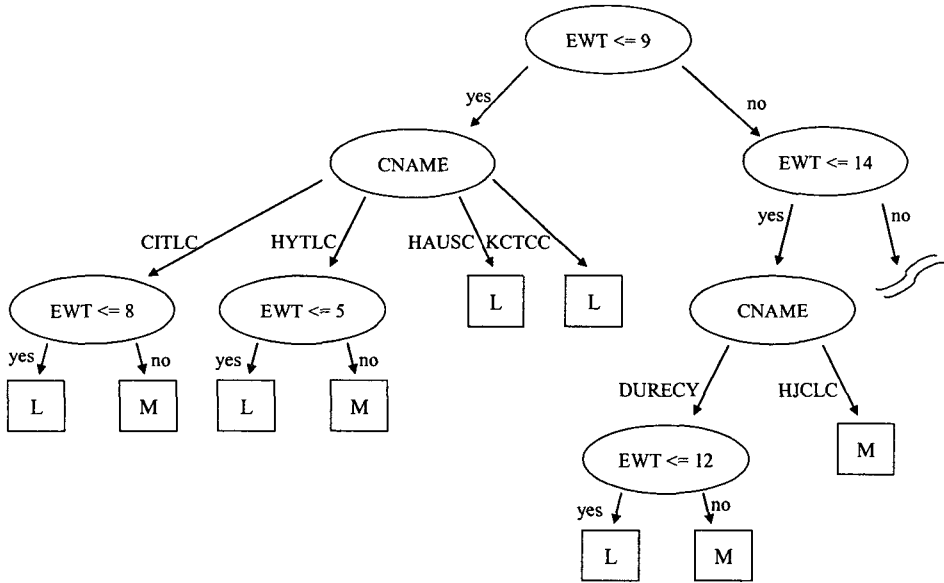
본 장에서는 제안한 방안의 효과를 확인해 보기 위하여 수행한 실험 결과를 정리하여 분석하였다. 수집한 5개월 분량의 수출 컨테이너 관련 데이터 중에서 먼저 터미널에 반입된 4개월 분량은 학습 시 훈련용으로 사용하였고, 나머지 1개월 분량은 평가를 위한 시험용으로 사용하였다. 20피트 컨테이너와 40피트 컨테이너에 대하여 각각 따로 실험하였으며, WEKA 데이터 마이닝(data mining) 소프트웨어를 이용하여 분류기를 생성하였다(Weka).

<표 13>은 학습을 이용하여 생성한 분류기들의 정확도를 비교하고 있다. 표에서 ‘운송사’는 현행 방법대로 운송사 무게 정보를 신뢰하여 무게그룹을 분류한 경우이며, 1R은 한 가지 컨테이너 정보만을 이용하여 규칙(rule)을 생성하는 OneR 학습 알고리즘을 적용한 경우이다(Holte, 1993). NB는 나이브 베이즈(Naïve Bayes) 학습 알고리즘을 사용한 경우로 수치 속성은 이산화(discretization)한 후 적용하였다(Friedman et al. 1997). DT는 결정 나무를 생성하는 J48(C4.5의 WEKA 버전) 학습 알고리즘을 적용한 경우이다(Quinlan, 1993). PART는 J48로 생성한 부분 결정 나무(partial decision tree)들을 규칙으로 변환하는 학습 알고리즘이다(Frank and Witten, 1998). 어떠한 컨테이너 관련 정보가 무게그룹 분류에 효과적인지 확인하기 위하여 원본 데이터를 모두 사용한 데이터1과 운송사 무게와 운송사 이름만을 사용한 데이터2 두 가지를 실험하였다.

<표 13> 학습을 이용한 무게그룹 분류의 정확도

무게그룹 분류 방법	무게그룹의 정확도			
	20피트		40피트	
	데이터1	데이터2	데이터1	데이터2
운송사	70.2%		64.5%	
NB	72.6%	72.9%	66.4%	66.2%
1R	72.8%	72.8%	64.7%	64.7%
DT	75.4%	75.6%	69.7%	67.9%
PART	74.3%	74.4%	69.7%	67.6%

표에서 알 수 있듯이 결정 나무를 사용하여 학습한 경우가 다른 학습 알고리즘을 적용한 경우보다 우수한 성능을 보였으며, 컨테이너 무게 정보 이외에 운송사 이름만 추가로 사용하더라도 분류



[그림 11] 생성된 결정 나무의 일부(40피트 컨테이너)

성능에 개선이 있음을 알 수 있다. 이는 운송사 별로 무게에 대한 기준이 꽤 다를 것이라는 가정을 뒷받침하는 결과이다.

[그림 11]은 데이터2를 이용하여 DT 방안으로 생성한 결정 나무의 일부분을 보이고 있는데 운송사에 따라 무게에 대한 기준이 다를 수 쉽게 파악할 수 있다. 그림에서 EWT와 CNAME은 <표 6>의 운송사 무게와 운송사명을 의미한다. 운송사 CITLC, HYTLC는 컨테이너 무게를 실제보다 가볍게 추정하는 경향이 있으며, 운송사 DURECY는 그 반대 경향을 가지고 있음을 살펴 볼 수 있다. 이 후 실험에서는 무게그룹 분류 시 정확도가 높고 컨테이너 크기에 관계없이 안정적인 성능을 보인 데이터1에 대하여 DT를 적용하여 실험하였다.

<표 13>은 생성한 분류기를 이용하여 장치하였을 때 예상되는 컨테이너 별 재취급 확률을 보이고 있다. 학습을 이용하여 무게그룹을 분류함으로써 40피트 컨테이너의 경우 재취급이 상당히 감소

한 결과를 볼 수 있으나, 20피트 컨테이너의 경우에는 무게그룹 분류의 정확도가 5% 가량 향상 되었음에도 불구하고 재취급은 오히려 미세하게 증가한 결과를 보이고 있다.

4장에서 소개한 비용행렬을 탐색하면서 학습하는 ‘학습(탐색)’의 경우 이러한 문제점이 일부 보완되어 재취급 발생 확률이 줄어들었다. ‘학습(탐색)’ 기법 적용 시 생성한 분류기를 평가하기 위하여 훈련 집합만으로 5분할 상호검증(5-fold cross-validation)을 사용하였으며([그림 10]의 evaluate_classifier에 해당), 탐색 시 현재 해에서 임의로 60개의 이웃해를 생성하고 개선이 없는 경우 탐색 진행을 중지하였다. 언덕 오르기 탐색 기법은 지역 해에 빠질 가능성이 있으므로 총 10번의 탐색을 수행하여 가장 좋은 평가를 받은 분류기를 결과로 사용하였다¹²⁾. <표 14>는 <표 13>을 운송사 무게

12) 5분할 상호검증으로 평가한 결과로 비교하였다.

그룹으로 장치하였을 경우와 비교하여 상대적인 재취급 확률의 감소 정도를 표현한 것으로 양수의 경우 재취급 발생이 감소한 것이며, 음수의 경우 재취급 발생이 증가한 것이다. 적게는 약 1%에서 많게는 약 11%까지 재취급 발생을 줄일 수 있었다. 20피트와 40피트 컨테이너 비율이 각각 절반씩 일 때¹³⁾ 전체적으로 <표 15>와 같은 재취급 확률의 상대적 감소 정도를 구할 수 있다. 운송사 무게를 이용하여 분류하는 경우에 비해 약 5~7% 재취급 발생을 줄일 수 있음을 예상할 수 있다.

<표 13> 장치 단수 별 컨테이너 재취급 확률

컨테이너 크기		개별 컨테이너의 재취급 확률			
		3단	4단	5단	6단
20 피트	운송사	0.150	0.187	0.215	0.236
	학습	0.151	0.188	0.216	0.237
	학습(탐색)	0.149	0.184	0.211	0.232
40 피트	운송사	0.207	0.263	0.307	0.343
	학습	0.184	0.236	0.278	0.313
	학습(탐색)	0.183	0.234	0.274	0.315

<표 14> 장치 단수 별 재취급 확률의 상대적 감소 정도

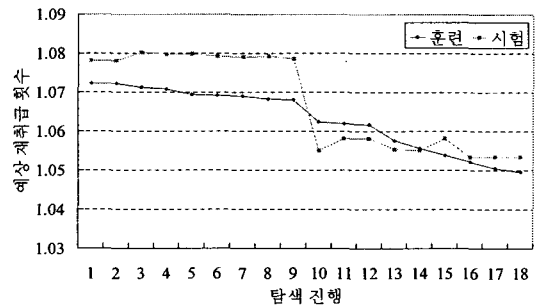
컨테이너 크기		상대적 재취급 비용			
		3단	4단	5단	6단
20 피트	운송사	0.0%	0.0%	0.0%	0.0%
	학습	-0.7%	-0.5%	-0.4%	-0.5%
	학습(탐색)	0.8%	1.4%	2.0%	1.7%
40 피트	운송사	0.0%	0.0%	0.0%	0.0%
	학습	11.1%	10.3%	9.4%	8.7%
	학습(탐색)	11.8%	11.0%	10.7%	8.1%

13) 실험에 사용하였던 10월분 데이터에는 20피트 컨테이너가 50.3%, 40피트 컨테이너는 49.7%를 차지하였다.

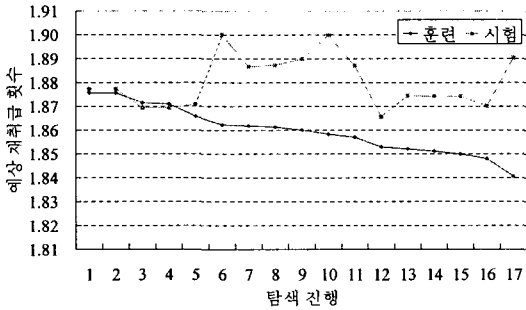
<표 15> 학습을 이용한 무게그룹 분류 시 장치 단수 별 재취급 확률의 상대적 감소 정도

장치 단수 별 재취급 확률의 변화			
3단	4단	5단	6단
7.2%	7.0%	7.1%	5.5%

<표 14>를 자세히 살펴보면 ‘학습(탐색)’ 방안의 경우 40피트 컨테이너를 6단으로 쌓을 때 ‘학습’ 방안에 비해 오히려 예상 재취급 발생 확률이 높아진 것을 알 수 있다. 이렇게 비용행렬 탐색에 의하여 개선이 이루어진 경우와 그렇지 못한 경우의 차이를 밝히기 위하여 탐색 과정을 분석하였다. [그림 12]는 탐색을 통하여 개선이 이루어진 20피트 컨테이너를 5단으로 장치하는 경우의 비용행렬 탐색 과정을 보이고 있다. 훈련 집합을 이용한 생성한 분류기의 평가값이 시험 집합의 평가값과 연관성이 있어 탐색이 효율적으로 이루어졌다. [그림 13]은 비용행렬 탐색으로 개선이 이루어지지 않은 40피트 컨테이너를 6단으로 장치하는 경우이다. 탐색이 진행됨에 따라 분류기에 대한 평가값은 개선되고 있지만, 시험 집합에 대한 평가값은 오히려 증가하는 학습에서의 과적합(overfitting) 현상이 발생하는 것을 확인할 수 있다.



[그림 12] 탐색 진행의 예 (20피트, 5단)



[그림 13] 탐색 진행의 예 (40피트, 6단)

<표 16>은 20피트 컨테이너를 5단으로 장치할 때 발견된 비용행렬로 무게그룹 L로 잘못 분류하는 경우인 첫 번째 행에 상대적으로 더 큰 비용이 명기되었다. 이는 무게그룹이 L로 분류된 컨테이너를 모으는 공간에 잘못 분류된 무게그룹 M 또는 H에 해당되는 컨테이너가 장치되는 경우, 적하 순서상 이들 컨테이너가 먼저 반출되어야 하므로 상단에 있는 무게그룹 L로 올바르게 분류된 여러 컨테이너들을 재취급하여야 하는 상황이 발생하기 때문이다. 반대로 두 번째와 세 번째 행에 해당되는 무게그룹 M 또는 H로 잘못 분류된 경우에는 잘못 분류된 컨테이너 자신만 재취급 될 가능성이 높기 때문에 상대적으로 낮은 비용이 부여되었다. 다른 관점으로 살펴보면 첫 번째 열의 무게그룹 L에 해당되는 컨테이너를 잘못 분류하는 경우에는 자신만 재취급 되는데 비해, 두 번째와 세 번째 열의 무게그룹 M과 H에 속하는 컨테이너는 잘못 분

<표 16> 발견된 비용행렬의 예(20피트, 5단)

예상한 무게그룹	선사 무게그룹		
	L	M	H
L	0.00	1.03	1.06
M	0.49	0.00	0.90
H	0.79	0.80	0.00

류되는 자신 이외의 다른 여러 컨테이너들을 재취급하여야 하는 상황이 될 가능성이 높기 때문에 상대적인 비용이 높게 매겨졌다.

<표 17>과 <표 18>은 각각 모든 잘못된 분류에 동일한 비용을 매기는 표준 비용행렬과 탐색으로 발견한 비용행렬을 이용하여 생성한 분류기의 혼돈행렬로 <표 16>의 비용행렬과 비교해 볼 때 상대적으로 비용이 높게 매겨진 항목에 대해서는 발생 비율이 줄었음을 확인할 수 있다¹⁴⁾.

<표 17> 표준 비용행렬을 사용하여 생성한 분류기의 혼돈행렬(20피트, 5단, 정확도 75.4%)

예상한 무게그룹	선사 무게그룹		
	L	M	H
L	13.6%	3.4%	1.9%
M	2.2%	9.8%	5.2%
H	4.1%	7.8%	52.1%

<표 18> 탐색으로 발견한 비용행렬을 사용하여 생성한 분류기의 혼돈행렬(20피트, 5단, 정확도 75.1%)

예상한 무게그룹	선사 무게그룹		
	L	M	H
L	11.6%	2.0%	1.0%
M	3.8%	11.2%	6.1%
H	4.2%	7.9%	52.1%

14) <표 16>의 비용행렬에서 무게그룹 H인 컨테이너를 M으로 잘못 분류할 때의 비용이 0.9로 전반적으로 그다지 작지 않음에도 <표 17>과 <표 18>을 비교해 보면 이렇게 분류된 컨테이너의 비율은 늘어난 것을 확인할 수 있다. 이는 무게그룹 H인 컨테이너의 수는 고정되어 있으며 이들 컨테이너를 무게그룹 L로 잘못 분류하였을 때는 더 높은 비용(1.06)이 발생하기 때문이다.

이상의 실험을 통하여 본 논문에서 제안하는 학습을 적용한 수출 컨테이너의 무게그룹 분류 방안의 유효성을 확인할 수 있었다.

6. 결론 및 향후 연구

본 논문에서는 컨테이너 터미널에서 운송사 무게 정보가 부정확하여 인해 발생하는 수출 컨테이너의 재취급 발생을 줄이기 위하여 기계학습을 적용하는 방안을 제시하였다. 이를 위하여 부정확한 무게 정보에 의하여 발생하는 재취급 횟수를 추정할 수 있는 모형을 소개하고, 비용행렬을 탐색함으로써 재취급 발생을 줄일 수 있는 분류기를 생성하는 방안을 제안하였다. 제안한 방안을 실험적으로 검증해 본 결과 재취급 발생을 5~7% 줄일 수 있음을 확인하였다.

운송사 무게 정보가 부정확한 경우 발생하는 재취급을 더욱 줄이기 위한 개선 방안으로 다음의 세 가지를 생각해 볼 수 있다. 먼저 본 논문에서는 하나의 스택에는 동일한 무게그룹으로 분류된 컨테이너들만 장치하는 것을 가정하였지만, 서로 다른 무게그룹간의 혼적을 허용한다면 재취급 발생을 더욱 줄일 수 있을 것이다. 이와 함께 특정 무게그룹 하나로 분류하지 않고 확률적으로 분류하는 분류기를 생성하고, 분류된 무게그룹 별 확률을 고려하여 장치하는 방안을 연구해 볼 필요가 있다. 또한 비용행렬 탐색 시 발생할 수 있는 과정합 현상을 해소할 수 있는 효과적인 분류기 평가를 적용한다면 재취급을 보다 적게 발생시키는 분류기를 생성할 수 있을 것이다.

참고문헌

- [1] 강재호, 류광렬, 김갑환, “장치장에서 베이 내 컨테이너의 효율적인 재정돈 방안,” *한국지능정보시스템학회 2004년 추계학술대회 논문집*, (2004a), 287-295.
- [2] 강재호, 오명섭, 류광렬, 김갑환, “반입 컨테이너 무게를 고려한 재취급 최소화 장치위치 결정 방안,” *한국지능정보시스템학회 2004년 추계학술대회 논문집*, (2004b), 271-278.
- [3] 강재호, 오명섭, 류광렬, 김갑환, “컨테이너 터미널 장치장에서 블록 내 이적을 위한 컨테이너 이동 순서 계획,” *한국항해항만학회 논문집 제29권 제1호*(2005), 83-90.
- [4] 류영욱, “컨테이너 터미널에서의 작업시간에 관한 연구,” 석사 학위 논문, 부산대학교 대학원 산업공학과, (1998).
- [5] 양지현, “재취급 최소화를 위한 혼적 결정 수리 모형 및 해법,” 석사 학위 논문, 부산대학교 대학원 산업공학과, (2003).
- [6] 오명섭, 강재호, 류광렬, 김갑환, “복수 크레인을 활용한 블록 내 컨테이너 이적 계획,” *한국항해항만학회 논문집*, 제29권 제5호 (2005), 447-445.
- [7] Domingos. P., “MetaCost: A general method for making classifiers cost-sensitive,” *Proceedings of International Conference on Knowledge Discovery and Data Mining, KDD'99*, (1999), 155-164.
- [8] Frank. E. and I. E. Witten, “Generating Accurate Rule Sets Without Global Optimization,” *Proceedings of the Fifteenth International Conference on Machine Learning*, (1998), 144-151.
- [9] Friedman. N., D. Geiger. D., and M. Goldszmidt, “Bayesian Network Classifier,” *Machine Learning*, Vol. 29(1997), 103-130.
- [10] Holte. R. C. “Very simple classification

- rules perform well on most commonly used datasets," *Machine Learning*, Vol. 11(1993), 63-91.
- [11] Kim, K. H. and J.-W. Bae, "Re-Marshaling Export Containers," *Computers and Industrial Engineering*, Vol. 35, No. 3-4(1998), 655-658.
- [12] Kim, K. H., Y. M. Park, and K. R. Ryu, "Deriving Decision Rules to Locate Export Containers in Container Yard," *European Journal of Operational Research*, Vol. 124(2000), 89-101.
- [13] Quinlan, J. R. *C4.5: Programs for Machine Learning*, San Mateo, CA, Morgan Kaufmann, (1993).
- [14] Weka3 - Data Mining with Open Source Machine Learning Software in Java <http://www.cs.waikato.ac.nz/~ml/weka/>
- [15] Witten, I. H. and E. Frank, *Data Mining-Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, (1999).

감사의 글

자료 수집에 도움을 주신 신선대 컨테이너 터미널 관계자분들께 지면을 빌어 감사의 마음을 전합니다.

Abstract

Learning a Classifier for Weight Grouping of Export Containers

Jaeho Kang* · Byoungho Kang* · Kwang Ryel Ryu* · Kap Hwan Kim*

Export containers in a container terminal are usually classified into a few weight groups and those belonging to the same group are placed together on a same stack. The reason for this stacking by weight groups is that it becomes easy to have the heavier containers be loaded onto a ship before the lighter ones, which is important for the balancing of the ship. However, since the weight information available at the time of container arrival is only an estimate, those belonging to different weight groups are often stored together on a same stack. This becomes the cause of extra moves, or rehandlings, of containers at the time of loading to fetch out the heavier containers placed under the lighter ones. In this paper, we use machine learning techniques to derive a classifier that can classify the containers into the weight groups with improved accuracy. We also show that a more useful classifier can be derived by applying a cost-sensitive learning technique, for which we introduce a scheme of searching for a good cost matrix. Simulation experiments have shown that our proposed method can reduce about 5~7% of rehandlings when compared to the traditional weight grouping method.

Key words : Container Weight Grouping, Rehandling, Cost-sensitive Learning

* Pusan National University

