대한산업공학회/한국경영과학회 2000 춘계공동학술대회 논문집. 2000 년 4 월 21-22 일. 경남대학교
Session PA10.3

혼동 행렬을 이용한 클러스터링 문제의 수리 계획적 접근

Mathematical Programming Application for Clustering Problems in Conjunction with Confusing Matrix

김영민, 최경현
한양대학교 산업공학과

Abstract

혼동 행렬 (confusion matrix)은 자극 또는 인식대상(데이터)에 대한 반응을 데이터화 함으로써 인식대상(데이터)의 특성분석을 통하여 복잡한 시스템을 효율적으로 통제, 관리하기 위한 분석기법에 사용된다. 클러스터링은 인식 시스템을 위한 기법으로서 다양한 분야에서 널리 활용되고 있다. 본 연구에서는, 혼동 행렬을 이용한 최적화 모델을 통하여 클러스터링(Clustering) 문제의 새로운 접근법을 제시한다. 최근 수리 계획 분야에서 클러스터링 분야에 대한 연구가 계속되고 있는데, 그러한 수리 모델과 혼동 행렬을 접목하여 새로운 모델을 제시한다.

1. 서론

급박하게 변해가는 현대 사회의 방대한 부산물인 데이터베이스는, 그 양과 질에서 나날이 발전해가며, 방대해지고 있다. 데이터베이스의 활용이 다양해지고, 규모가 커짐에 따라 효율적인 사용이 중요시되고 있는데, 작업에 가져오는 이득을 위해서 그런 데이터를 일반화하는 프로세스 모델이 필요하게 되었다. 즉, 데이터를 어떻게 이해하느냐가 중요한 관심사가 된 것이다.

현재, 방대한 데이터를 효율적으로 다루어, 여러 가지 의미 있는 정보를 효과적으로 추출해 내는 것에 많은 관심이 집중되고 있다. 그러나, 그러한 기술에 대한 필요성에 비해 아직 체계화된 방법들이 많이 정립되어 있지는 않다. 데이터로부터 정보를 추출하는 것에 대한 필요성이 점점 커짐에 따라, 각 응용 분야에서 각자의 특수한 조건에 맞추어 정보를 얻어내는 기술이 많이 개발 되었지만, 이를 일반화하는 모델은 그리 많지 않다는 것이다.

데이터를 이해하기 위한 체계화된 기술로서, 클러스터링(clustering), 분류(classification) 등을 꼽을 수 있다. 이는 데이터를 분류하고, 특성을 찾아내기 위한 기술이며, 다양한 적용 분야에서 여러 가지 방식으로 응용되어 쓰이고 있다.

간단히 말해, 원하는 특성을 찾아내기 위해 그룹을 나누는 것이라 할 수 있는 클러스터링은 인지 시스템에서 가장 기초적인 이론이자 기술이라고 할 수 있다. 인지 시스템과 같은 위상을 갖고 있는 데이터 마이닝과 machine learning 등에서도 클러스터링은 중요한 부분을 차지하고 있다.

본 논문에서는 인지 시스템의 자극-반응 행렬에서 사용하는 혼동 지수를 클러스터링에 적용해 보고자 한다. 먼저, 인지 및 인간 공학에서 사용하는 자극-반응 행렬을 이용한 혼동 모델에서 클러스의 의미를 분석해 보겠다. 기존 모델에 일반적인 클러스터링의 의미를 포함시켜 발전된 모델을 제시하고, 이러한 모델에서 다시 일반적인 클러스터링을 위한 아이디어를 얻어낸다. 이렇게 제안하는 모델은 수리 계획적 측면에서 접근한 것이며, 기존 모델과 비교하여 다양한 해석이 가능함을 보여주겠다. 여기서 제안하는 새로운 클러스터링 모델은 혼동 지수의 개념을 적용한 것이다. 본 논문에서 제안된 정형화된 수리 모델을 통해, 그 해법에 있어서도 보다 효율적임을 확인할 수 있을 것이다.

2. 본론

클러스터링과 혼동 행렬 및 혼동 지수에 대한 의미를 명확히 하고, 이를 통해 클러스터링 모델을 새로운 방식으로 접근할 수 있음을 보인다. 수리계획 분야에서 Bradley et al.(1999)이 제안한 convex combination 을 사용한 모델을 응용하도록 하겠다.

2.1 클러스터링과 혼동 행렬

클러스터링(clustering)이란 주어진 m 개의 점 또는 데이터를 몇 개의 클러스터(cluster) 또는 집합으로 묶음으로써, 각각의 클러스터 또는 집합에 포함된 점(데이터)들의 특성을 찾아내는 것이다 클러스터링의 목적은 데이터를 비슷한 성질을 갖는 점들의 집합으로 그룹화 하는 것이며, 클러스터링의 목표는 각각의 클러스터의 독특한 특성이 그 내부에 포함된 점들로서 잘 표현되도록 하자는 것이다.

일반적인 클러스터링 알고리즘은 두 단계를 거친다. 첫 번째 단계에서 클러스터의 개수를 정하고, 두 번째 단계에서 그 정해진 개수를 갖고 최적의 클러스터링을 하는 것이다. 이러한 알고리즘을

반복함으로써 원하는 클러스터링을 할 수 있게 된다. 클러스터링은 다양한 분야에 응용될 수 있으므로, 그 적용되는 분야에 따라 서로 다른 여러 가지 제약이나 조건들을 포함하는 모델을 만들어 볼 수 있을 것이다.

혼동 행렬이란, 인지 시스템의 설계를 위해 제안된 것으로, 자극-반응 부합성 위배와 집단화 및 기능적 군집화 결여, 반응 혼동의 문제를 최소화 하기 위해 만들어졌다. Kantowitz and Sorkin(1983)과 Sanders and McCormick(1982)는 Green and Pew(1978)의 행렬을 인용하여 혼동 행렬에 대한 분석 내용을 공통적으로 다루고 있다. 이러한 혼동 행렬에 관한 연구의 기본 목적은 자극-반응이 효과적으로 부합하는 쌍을 찾아내어 시스템을 재설계함으로써 인식도와 차별성을 향상시키는 데 있다. 인지 측면에서 이를 조금 더 발전시키면, 혼동을 최대한 피하는 방법으로 인지 대상을 몇 개의 클러스터로 나눈다는 목적을 갖는 모델을 만들어 볼 수 있다. 또한 이 혼동이라는 것을 어떻게 정의하느냐에 따라 일반적인 클러스터링으로 확대될 수도 있을 것이다.

수리계획 측면에서 접근한 클러스터링은, 주어진 데이터를 n 차원상의 점이라고 생각하고, 정해진 클러스터의 개수에 맞게 클러스터 중심을 찾는 것으로 시작한다. 주어진 m 개의 점을 n 차원의 실수 공간(R^n)에서 k 개의 cluster 로 할당하는 문제는, 각각의 점에서 가장 가까운 center 까지의 거리들의 합이 최소화 되도록 R^n 에서 k 개의 center 를 정하는 것으로 생각할 수 있다.

2.2 혼동행렬을 사용한 자극 할당 문제의 수리모델

자극-반응 행렬은 인간공학, 인지공학, 그리고 여러 응용심리 분야에서 자극에 따른 반응 상태를 분석하기 위하여 폭 넓게 사용되고 있다. 이는 다양한 종류의 자극과 반응에 대한 혼동의 특성 분석을 통하여 반응이 유사한 자극 또는 상이한 자극의 종류를 찾는 데 사용된다. 이러한 의미에서 자극-반응 행렬을 혼동 행렬이라고도 한다. Theise (1989)는 이러한 혼동 행렬을 사용하여 최적화 모델을 만들었는데, 그 중 첫번째 모델은 다음과 같다.

$$\min \sum_{i=1}^{m} \sum_{j=i+1}^{m} \mathbf{a}_{ij}\mathbf{x}_i\mathbf{y}_j + \sum_{i=1}^{m} \mathbf{b}_i\mathbf{x}_i \tag{1}$$

$$\text{subject to} \sum_{i=1}^{m} \mathbf{x}_i = \text{s}$$

$$\mathbf{x}_i \in \{0,1\}, \quad \text{for all } i = 1,\ldots,m$$

이 모델은 주어진 여러 가지의 자극 중에서 원하는 개수의 자극 집합을 형성하는 문제이며, 주어진 자극-반응 행렬과, 주어진 원하는 그룹에 포함된 자극의 개수 s 를 만족하는 상호 최소 혼동 지수를 가진 자극을 찾는 모델이다. 즉, 이 모델을 통해 선택된 자극의 집합은 서로 혼동할 우려가 최소화 된 자극들의 집합이라는 것이다. $\mathbf{x}_i$ 는 i 번째 자극을 의미하는데 만약 $\mathbf{x}_i$ 가 선택되면 1, 그렇지 않으면 0 의 값을 갖는다. 또 $\mathbf{a}_{ij}$ 는 행렬의 원소를 나타내고, $\mathbf{b}_i$ 는 i 번째 자극을 인식 못한 경우(마지막 열)의 수를 나타낸다. 이 모델에서의 목적함수는 i 번째 자극과 j 번째 자극이 동시에 1 인 경우(원하는 집합에 포함되는 경우) 상호간의 혼동 지수 $\mathbf{a}_{ij}$ (i 번째 자극을 j 로 인식하는 경우와 j 번째 자극을 i 로 인식하는 경우의 곱) 값의 합과, 자극을 인식하지 못한 경우의 가중치 $\mathbf{b}_i$ 를 합한 값을 최소화하는 것이다.

이 모델에서는 원하는 집합을 하나밖에 구하지 못한다는 단점을 갖는다. 실제 응용 문제에서는 하나의 집합만을 구하는 경우는 그리 많지 않다. 따라서 이 모델을, 선택되는 자극들 상호간의 혼동을 최소화하는 여러 개의 집합을 구하는 문제로 발전시킨다면, 여러 개의 클러스터를 갖는 클러스터링 문제로 확장할 수 있을 것이다. 또한 계수 $\mathbf{a}_{ij}$ 와 $\mathbf{b}_i$ 를 어떻게 정의하느냐에 따라 일반적인 클러스터링 문제로 확장 가능하다.

먼저, 식 (1)에서 여러 개의 집합으로 나누는 문제로 확장해 보자. m 개의 자극을 k 개의 상호 최소 혼동 지수를 갖는 부분 집합으로 나누는 모델을 생각해 본다. 이 문제의 모델을 설정하기 위하여 m 을 자극(혹은 인지 대상)의 개수라 두고, k 를 원하는 부분집합의 개수, S_k^L 와 S_k^U 를 각각 부분집합 k 에 들어갈 수 있는 자극(인지 대상)의 개수의 하한과 상한이라 둔다. 모든 i(1≤i≤m)와 l(1≤l≤k)에 대하여 0-1 변수 $\mathbf{x}_{il}$ 을

$$\mathbf{x}_{il} = \begin{cases} 1, & \text{i 번째 자극이 l 번째 부분집합에 속함} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

이렇게 정의하면, 자극 할당 문제는 다음과 같이 바뀐다.

$$\min \left(\max_{l \le l \le k} \sum_{i=1}^{m} \sum_{j=1}^{m} \mathbf{c}_{ij}\mathbf{x}_{il}\mathbf{x}_{jl} \right) \tag{2}$$

$$\text{subject to} \sum_{l=1}^{k} \mathbf{x}_{il}\mathbf{x}_{jl} = 1, \ \forall i = 1,\ldots,m \tag{3}$$

$$S_l^L \le \sum_{i=1}^{m} \mathbf{x}_{il} \le S_l^U, \ \forall l = 1,..k$$

$$\mathbf{x}_{il} \in \{0,1\}, \ \forall i = 1,\ldots,m, \ \forall l = 1,\ldots,k$$

식 (2)에서 max 를 취하는 대신, 평균을 취하는 방법 등을 사용해서 식을 약간 변형할 수도 있다. 어떤 문제에 적용하느냐에 따라 목적식이 약간씩 바뀔 수 있을 것이다.

2.3 일반적인 클러스터링 문제로의 적용

앞서 설명한 바와 같이, 일반적인 클러스터링은 클러스터의 개수를 미리 정하고, 각 클러스터의 중심이 되는 점을 구한 후, 그 중심점들에 최적으로 점들을 할당한다. 이를 식으로 표현하면 다음과 같이 나타낼 수 있다. 아래 모델은 그 과정을 동시에 고려했지만, 클러스터의 중심을 구하는 것과 점을 할당하는 것은 대개 순차적으로 일어난다.

$$\min_{\{c^1,\Lambda,c^k\}} \left\{ \sum_{i=1}^{m} \min_{1 \le j \le k} \| \mathbf{x}^i - \mathbf{c}^j \| \right\} \tag{4}$$

식 (4)는 주어진 점에서 가장 가까운 중심점(Center)을 찾아 그 중심에 해당되는 클러스터에 점을 할당한다. 즉, 클러스터에 할당하는 기준은 점과 중심 사이의 거리가 되는 것이다. 그리고, 중심을 정하는 것과 주어진 점을 할당하는 것 모두 변수로 둘 수 있다. 비록, 나뉘어진 점들은 중심과만 비교된 것이지만, 같은 클러스터로 할당된 점들은 서로 비슷한 특성을 갖고 있다고 생각한다. 그 이유는 각각의 점들의 위치 기준이 되는 좌표가 결국 그 점들의 특성을 기준으로 한 것이기 때문이다. 점들의 좌표가 가깝다는 것은 그 점들의 유사성을 나타낸다고 생각해도 무방하다.

이 문제는 convex 및 nonconvex 에 모두 해당되지 않는 매우 복잡한 문제이다. 따라서, Bradley et al.(1999)은 수리 계획의 문제에서의 dual 의 특성을 활용하여 이 문제를 다음과 같이 변환시켰다.

$$\min_{\mathbf{c},\mathbf{t}} \sum_{i=1}^{m}\sum_{j=1}^{k} \mathbf{t}_{ij} \| \mathbf{x}^i - \mathbf{c}^j \| \tag{5}$$

$$\text{subject to } \sum_{j=1}^{k} \mathbf{t}_{ij} = 1, \ \ \forall i = 1,\dots,m \tag{6}$$

$$\mathbf{t}_{ij} \ge 0, \ \ \forall i = 1,\dots,m, \ \forall j = 1,\dots,k$$

식 (5)는 각 점과 가장 가까운 중심을 선택하는 과정이 각 점과 중심들 사이의 convex combination 으로 대치된 것임을 보여준다. 식 (6)은 convex rate 의 합이 1이 됨을 나타낸다.

위의 모델에 앞에서 설명한 혼동 지수의 개념을 넣어보자. 이제, 점을 클러스터에 할당하는 과정에서 그 기준이 되는 지표로 각 점과 가장 가까운 중심을 선택하는 것이 아니라, 새로운 것을 선택하게 된다. 즉, 혼동 지수의 개념을 넣게 되면, 임의의 어떤 점을 하나의 클러스터에 할당하는 기준은, 그 중심이 되는 점과의 혼동을 가능한 최소로 할 수 있도록 하는 것이 된다는 것이다. 문제에 따라서 반대로, 혼동을 최대화 하도록 바꿀 수도 있다.

식 (5)에서 $\| \mathbf{x}^i - \mathbf{c}^j \|$ 대신 점 i 와 j 의 혼동 지수를 나타내는 $\mathbf{a}_{ij}$ 를 포함한 term 을 넣어준다. 여기서 혼동 지수는 주어진 두 점 사이에서 정의된다. 따라서 이를 이용한 모델을 만들기 위해서는 중심으로 채택되는 점도 주어진 점들 중 하나로 하는 것이 합당하다. 또한 선택되는 클러스터의 수는 k 개로 정해져 있으므로 그러한 제약식도 생각해 주어야 한다. 여기에 식 (2)에서 제안한 이진 변수의 개념이 들어가게 된다. 이 모델의 장점 중 하나는 식 (5)로 표현된 모델이 두 단계를 거쳐야지만 답을 구할 수 있게 되어 있는 반면, 여기선 중심을 구하는 과정이 모델에 모두 포함되어 있다는 것이다. 또한 k 가 식에 포함됨으로써, k 를 변수로 두는 문제로 응용해 볼 수 있다는 장점이 있다.

지금까지 설명한 내용을 수식으로 표현해 보자.

$$\min_{\mathbf{y},\mathbf{t}} \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbf{a}_{ij}\mathbf{t}_{ij}\mathbf{y}_j \tag{7}$$

$$\text{subject to } \sum_{j=1}^{m} \mathbf{t}_{ij}\mathbf{y}_j = 1, \ \ \forall i = 1,\dots,m \tag{8}$$

$$\sum_{j=1}^{m} \mathbf{y}_j = k \tag{9}$$

$$0 \le \mathbf{t}_{ij} \le 1, \ \ \forall i = 1,\dots,m, \ \forall j = 1,\dots,m \tag{10}$$

$$\mathbf{y}_j \in \{0,1\}, \ \ \forall j = 1,\dots,m \tag{11}$$

결정 변수 $\mathbf{y}_j$ 는 이진 변수로서, 점 $\mathbf{x}^j$ 가 클러스터의 중심으로 선택되면 1, 그렇지 않으면 0 의 값을 가진다. 또한 결정 변수 $\mathbf{t}_{ij}$ 는 convex combination 의 가중치이다. 이러한 두 개의 변수와 혼동 지수 $\mathbf{a}_{ij}$ 를 사용하여 식 (5)를 식 (7)로 바꾸어 표현할 수 있다. 식 (8)은 convex rate 의 합이 1 이 된다는 식이며, $\mathbf{y}_j$ 가 1 일 때 의미를 갖게 되므로 $\mathbf{y}_j$ 를 각 항에 곱해주었다. 식 (9)는 중심으로 선택되는 클러스터의 개수를 k 로 제한하는 식이다.

기본적인 모델을 이렇게 만들어 볼 수 있지만, 수리계획을 이용한 기존 클러스터링 모델에 혼동 개념을 적용했을 때 약간 문제가 생김을 알 수 있다. 즉, 본래 혼동 지수는 각 클러스터에 속하는 점들 상호 관계를 모두 생각해 주는 것이지만, 이 모델에서는 중심과만 비교하는 것이 된다. 그 결과 각 클러스터로 할당된 점들 상호간의 혼동 지수는 고려하지 않게 된다. 그러한 문제를 원천 봉쇄하기 위해 이 모델에서 혼동 지수 대신 점들의 좌표를 이용한 다른 지표를 넣어 줄 수 있다. 기존의 문제 정의에서와 같이 두 점 사이의 거리로 해도 된다. 그렇게 되면 지금까지 설명한 혼동 지수는, 직접적으로 사용하게 되지 않으므로 무의미해 보일 수 있다. 그러나, 이 모델이 나오게 된 배경을 뒷받침해주므로, 의미는 그대로 갖고 있게 된다. 따라서, 식 (7)은 다음과 같은 식으로 바꿀 수 있다.

$$\min_{\mathbf{y},\mathbf{t}} \sum_{i=1}^{m}\sum_{j=1}^{m} \| \mathbf{x}^i - \mathbf{x}^j \| \mathbf{t}_{ij}\mathbf{y}_j \tag{12}$$

2.4 인지 시스템에 새로운 클러스터링 모델 적용

이제 다시 인지 시스템으로 돌아가보자. 일반적인 클러스터링은 서로 비슷한 점들끼리 묶는다는 특성이 있지만, 혼동을 최소화 하는 것이 목적일 때는 서로 다른 성질을 가져서 구별하기 쉬운 점들끼리 묶게 된다. 따라서 혼동 지수 $\mathbf{a}_{ij}$ 를 그대로 쓰는 인지 시스템의 클러스터링 모델에서는 앞서 얘기한 바와 같이 중심과만 비교해서 점들을 클러스터링 해서는 않된다. 만약 그렇게 한다면, 각 클러스터에 모인 점들 각각은 중심과는 잘 구분될 지 몰라도 서로간에는 어떻게 될 지 보증할 수가 없기 때문이다.

여기서 인지 시스템의 문제 정의를 한 번 다르게 해보자. 만약, 클러스터링의 목적이 각 클러스터에 속한 점들 상호간의 혼동을 줄이도록 하는 것이 아니라, 기존의 클러스터링과 마찬가지로 각 클러스터에 속하는 점들의 속성이 비슷하도록 하는 것이라면, 혼동 지수를 최대화 하는 모델을 생각해 볼 수 있다. 그렇게 되면, 각 클러스터에는 비슷한 성격을 가진 점들이 모이게 되며, 그러한 특성을 유용하게 쓸 수 있는 인지 시스템 문제에 적용할

수 있다. 여기서 각 클러스터 사이의 혼동 지수를 최소화하는 (클러스터끼리의 구분을 더 확실하게 하기 위함) term 을 목적식에 추가하면 좀 더 정확한 모델을 만들 수 있을 것이다. 이를 적용한 모델은 다음과 같다.

$$\max_{\mathbf{y},\mathbf{t}} \left\{ \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\mathbf{a}_{ij}\mathbf{t}_{ij}\mathbf{y}_j + \frac{1}{\mathbf{a}_{ij}}\mathbf{y}_i\mathbf{y}_j \right) \right\} \qquad (13)$$

$$\text{subject to } \sum_{j=1}^{m}\mathbf{t}_{ij}\mathbf{y}_j = 1, \quad \forall i = 1,\dots,m$$

$$\sum_{j=1}^{m}\mathbf{y}_j = k$$

$$0 \le \mathbf{t}_{ij} \le 1, \quad \forall i = 1,\dots,m, \ \forall j = 1,\dots,m$$

$$\mathbf{y}_j \in \{0,1\}, \quad \forall j = 1,\dots,m$$

제약식은 목적식 (7)에서와 마찬가지로 동일하다. 식 (13)은, 각 클러스터 내의 점들의 혼동 지수를 최대화 하면서, 클러스터 서로를 잘 구분할 수 있도록 클러스터 간의 혼동 지수는 최소화하는 모델이다. 이 식에서는 $(1/\mathbf{a}_{ij})\mathbf{y}_i\mathbf{y}_j$ 부분이 클러스터 간의 혼동 지수를 최소화 하는 부분이지만, 계수 $1/\mathbf{a}_{ij}$ 을 혼동 지수 $\mathbf{a}_{ij}$를 사용한 다른 식으로도 바꾸어 모델을 조금 변형할 수 있다. 또한 각 term 에 가중치를 줌으로써, 중요도를 달리하여 다양한 응용 문제에 맞추어 적용해 볼 수 있다.

3. 결론

지금까지 혼동 지수를 사용한 인지 시스템, 수리 계획 분야에서 접근한 클러스터링, 이 둘의 조합으로 만든 새로운 클러스터링 모델에 대해 살펴보았다.

혼동 행렬을 이용하여 혼동을 최소화 하는 여러 개의 클러스터를 만드는 문제는 인지 및 인간공학의 다양한 분야에 적용해 볼 수 있을 것이다. 쉬운 예로, 키보드 배치나 버튼 위치 설정 등을 생각해 볼 수 있다. 혼동 지수를 사용한 모델에서 아이디어를 얻어 만든 클러스터링 모델, 식 (12), (8)- (11)은 클러스터링 과정을 한 단계로 표현 가능하다는 것을 보여준다. 이 식은 선형화를 통해 보다 쉽게 풀 수 있을 것이다. 마지막으로 제안한 식 (13)은 다양한 방법으로 응용 가능하다. 목적식에 항이 추가됨으로써 이전보다 복잡한 식이 되었지만 적절한 식 변형을 통해 쉬운 문제로 바꿀 수 있을 것이다.

본 논문에서는 클러스터링 문제의 새로운 모델을 제시했다. 앞으로 추후 연구 과제로는 좀 더 현실적인 상황에 맞춰 모델을 응용 하여 적용해 보는 것과, 이 모델에 알맞은 알고리즘을 개발하는 것이다. Lagrangian dual 과 대규모 최적화 기법을 이에 적용해 볼 수 있으리라 생각한다.

4. 참고 문헌

Bennett, K. P., Mangasarian, O. L., Robust linear programming discrimination of two linearly inseparable sets, *Optimization Method and Software*, 1, 23-34, 1992,

Bradley, P. S., Fayyad, U. M., Mangasarian, O. L., Mathematical Programming for Data Mining: Formulations and Chalenges., *INFORMS Journal on Computing*, 11, 217-238, 1999.

Bradley P. S., Mangasarian O. L., Street W. N., Clustering via Concave Minimization., *Advances in Neural Information Processing Systems 9*, MIT Press, Cambridge 368-374, 1997.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., The KDD process for extracting useful knowledge fron volumes of data., *Communications of the ACM*, 39, 27-34, 1996.

Fisher. D., Knowledge acquiisition via incremental conceptual clustering., *Machine Learning*, 2, 139-172, 1987.

Green, P., and Pew, R. W., Evaluation pictographic symbols: An automotive application., *Human factors*, 20, 103-114, 1978.

McCormick, E. J., and Sanders, M. S., *Human factors in engineering and design*. New York: McGrawHill, 1982.

Morgan, B. J. T., Cluster analyses of two acoustic confusion matrices., *Perception and Psychophysics*, 13, 13-24., 1973.

Moore, T. G., Tactile and kinaesthetic aspects of push-buttons., *Applied Ergonomics*, 5, 66-71., 1974.

Theise, E. S., Finding a Subset of Stimulus-Response Pairs with Minimum Total Confusion: A Binary Integer Programming Approach., *Human factors*, 31, 291-305, 1989.

Townsend, J. T., Theoretical analysis of an alphabetic confusion matrix., *Perception and Psychophysics*, 9, 40-50., 1971.