

미 경영대학원 입학 희망자를 위한 학교 선별 시스템

윤기범, 구분화, 김남훈, 고한석
고려대학교 전자공학과

School Selection System for MBA Applicants

Gi-Bum Yoon, Bon-Hwa Ku, Nam-Hoon Kim, Hanseok Ko
Dept. of Electronics Engineering, Korea University
{gbyoon, bhku, nhkim}@ispl.korea.ac.kr, hsko@korea.ac.kr

요약

본 논문에서는 판단 보조 시스템으로써의 기능을 수행하는 Artificial Neural Network(ANN) 시스템의 응용과 설계를 수행하였다. ANN 시스템의 응용에 있어 실질적으로 활용이 가능한 미 경영대학원 진학 희망자들을 위한 학교 선별 시스템을 구현하였다. 미국 소재의 대부분의 경영대학원에서는 입학 지원자의 평가 항목으로써 대학평점(GPA), TOEFL(Test Of English as Foreign Language), GMAT(Graduate Management Admission Test), Essay 그리고 실무 경력을 요구한다. 본 논문에서 소개하는 학교 선별 시스템에서는 이러한 평가 항목 중 객관적인 평가가 가능하고 수치화 할 수 있는 세가지 요소만을 고려했다. 즉, Essay와 실무 경력과 같은 평가 항목은 각 학교별 다소 주관적인 평가가 이루어지기 때문에 배제하고 보다 객관적이라고 할 수 있는 GPA, TOEFL, GMAT만을 평가 항목으로 고려했다.

일반적으로 경영대학원에서는 최소의 TOEFL 점수를 요구하고 있다는 점을 고려해 TOEFL 점수에 우선 순위를 두고 각 학교를 크게 4개의 그룹으로 나누어 4개의 각기 다른 Network를 설계했다. 최소 5년 이상의 유학 상담 경험이 있는 현직 전문가들을 통해 본 논문에서 소개하는 시스템의 성능을 검증 받았다.

I. 서론

지금까지 미국 소재 경영대학원 진학을 위해 알려져 있는 다양한 정보들은 양이 너무 방대한 반면, 대부분 불확실하다. 따라서 미 경영대학원 진학 희망자들에게 있어 지원자의 성적에 가장 적합한 학교 선별을 위한 정형화된 규칙을 만들어 낸다는 것은 매우 어려운 일이다. 그러나 최근의 판단 도출을 위한 인공지능

기법 즉, Artificial Neural Network, Fuzzy Set Theory 등의 발달로 이러한 불확실성을 떠는 정보의 일반화가 가능해졌다. 본 논문에서는 불확실한 다양한 정보로부터 일반화된 규칙을 얻어내기 위한 방법으로 Fuzzy Set Theory와 Artificial Neural Network를 적용한다.

지금까지 다양한 종류의 ANN의 응용 시스템에 관한 연구가 진행되어져 왔다. 물체의 분류, 로봇 제어, 음성 인식, 데이터 압축등과 같은 연구들이 그것이다. 그러나 이러한 연구들은 인식률이라는 확실한 평가 요소를 갖는다. 다시 말해 기존의 방법 대신 ANN이라는 새로운 방법을 적용해 그 성능의 향상을 꾀하고자 하는 접근 방법이었다. 그러나 본 논문에서는 기존 연구와는 다른 즉, 이미 존재하는 시스템에 새로운 방법으로 접근하는 것이 아닌 다른 사람들이 생각하지 못한 학교 선별 시스템이라는 새로운 시스템을 구현했다.

미 경영대학원 지원자를 위한 ANN 시스템을 구현하기 위해서 훈련 데이터인 최신의 학교 입학 정보를 다수의 인터넷 웹 사이트를 통해 얻었다. 대부분의 경영대학원에서 GPA, TOEFL, GMAT, Essay, 그리고 실무 경력의 제출을 요구했다. 본 논문에서는 이러한 요구 사항 중 객관적이고 수치화할 수 있는 GPA, TOEFL, GMAT 세가지 점수만을 고려한다.

본 논문에서는 우선 인터넷을 통해 수집한 학교 명단을 Fuzzy clustering 알고리즘을 이용하여 그룹화하고 이렇게 그룹화 된 학교 명단들을 이용해 ANN 시스템을 훈련시킨다. 이러한 학교 선별 시스템의 전체적인 개요를 그림 1에 나타내었다. 본 논문에서 소개하는 시스템은 그림에 나타낸 것 과 같이 입력으로 경영대학원 진학 희망자의 GPA, TOEFL, GMAT 점수를 받게 되고, 훈련된 ANN System의 출력으로 지원자의

점수에 가장 적합한 경영대학원 학교 리스트를 보여주게 된다.

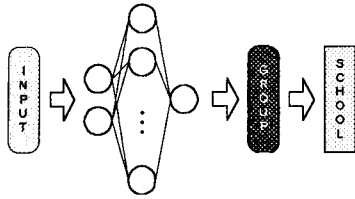


그림 1. 시스템 개요도

본 논문은 다음과 같이 구성된다. 다음의 2 장에서는 미국 소재 경영대학원의 과정과 입력 정보로 사용된 각 시험에 대해서 간략하게 소개하고, 3 장에서는 Fuzzy clustering 알고리즘과 Backpropagation 알고리즘을 소개한다. 그리고 새롭게 설계된 학교 선별 시스템에 대한 구체적인 설명은 4 장에서 이루어지고, 마지막으로 5 장과 6 장에서는 각각 본 논문에서 소개하는 시스템의 성능 평가와 결론에 대해 이야기한다.

II. Preliminary

2.1 경영학 석사 (MBA)

MBA(Master of Business Administration)는 경영 전문학 석사 학위 과정 프로그램이다. 상위 경영 관리자로서의 이론적인 지식뿐 아니라 실무적인 지식 습득이 목표이며 수업은 경영 전반에 걸친 내용과 실 사례 및 경험 그리고 대인 관계에 관련된 직접 사용할 수 있는 가장 실용적인 내용들을 다룬다. MBA 과정은 현재의 글로벌 시장에 적합한 폭 넓고도 실용적인 주제를 다루는 전문 경영인 프로그램이다.

2.2 GMAT

GMAT(Graduate Management Admission Test)는 미 경영대학원에 입학하기 위해서는 반드시 응시해야 하는 필수 시험이다. 응시자의 일반적인 문법, 수학, 분석에 입각한 작문 능력을 평가하기 위한 시험으로 경영에 관련한 지식을 평가하기 위한 시험은 아니다. GMAT는 응시자의 대학원 수업의 기본적 수학 능력을 평가하는 시험으로 GPA(대학 성적)과 더불어 학교로 하여금 MBA 과정 지원자의 학업 성취 정도를 가늠할 수 있게 하는 중요한 자료이다.

2.3 TOEFL

TOEFL(Test of English as Foreign Language)의 목적은 영어가 모국어가 아닌 사람들을 대상으로 그들의 영어에 대한 숙달 정도를 평가하기 위한 시험이다. 북미 지역 영어를 기준으로 격식화된 영어에 대한 이해와 활용도를 평가하는 시험이다. TOEFL은 미국 및

캐나다 소재 대학 또는 대학원에서의 수학을 희망하는 외국 학생들이 반드시 치루어야 할 시험이다.

III. Clustering and Learning

3.1 Fuzzy c-means Clustering

어떤 집합을 적합한 공통된 특성에 따라 몇 개의 그룹으로 나누는 것을 clustering 이라 한다. Fuzzy c-mean clustering 이 고전적인 Hard clustering 과 구별되는 가장 큰 차이점은 집합의 어느 한 원소가 오직 하나의 그룹에만 속하느냐의 여부이다. Fuzzy c-means clustering 의 경우 한 원소가 여러 개의 그룹에 속할 수 있음을 인정하고 그 소속 정도를 멤버십 함수를 통해 표현하게 된다. Fuzzy c-means (FCM) 알고리즘[1] 은 Fuzzy clustering 의 방법들 중 가장 폭 넓게 쓰이는 방법으로 Fuzzy c-partition 개념을 기반으로 하고 있다. Fuzzy clustering 알고리즘은 Ruspini[2]에 의해 소개 되었으며 간략하게 다음과 같이 정리될 수 있다.

어떤 임의의 집합에 $X = \{x_1, \dots, x_n\}$ 가 주어진 데이터의 한 집합이라고 하면 각 원소인 $x_k (k = 1, \dots, n)$ 는 R^p 차원의 벡터이다. 여기서 U_{cn} 을 $c \times n$ 매트릭스로, c 를 $2 \leq c < n$ 범위의 정수로 정의하면 X 에 대한 Fuzzy c-partition space 는 다음과 같다.

$$M_{fcn} = \{U \in U_{cn} : u_{ik} \in [0,1],$$

$$\sum_{i=1}^c u_{ik} = 1, 0 < \sum_{k=1}^n u_{ik} < n\} \quad (1)$$

여기서 u_{ik} 는 x_k 의 그룹 $i (i = 1, \dots, c)$ 에 대한 멤버십 값이다.

FCM 알고리즘의 목표는 최적화된 Fuzzy c-partition 과 그에 상응하는 아래의 목적함수(식(2))를 최소화 시키는 prototype 을 찾는 것이다.

$$J_m(U, V; X) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2 \quad (2)$$

식(2)에서 $V = (v_1, v_2, \dots, v_c)$ 는 알려지지 않은 클러스터의 중심(prototype) $v_i \in R^p$ 의 매트릭스이다. $\|\cdot\|$ 은 유클리디안 norm 이고, $[1, \infty)$ 범위의 가중치 지수 m 은 멤버십 값에 영향을 미치는 상수이다.

식(1)에 정의된 제한요건을 만족시키면서 목적함수 J_m 을 최소화 시키기 위한 FCM 알고리즘은 선택적 최소화 알고리즘으로 다음과 같이 정의되어 진다. 우선 작은 양의 상수값 c, m, ϵ 를 갖는 벡터를 선택한다. 그리고 랜덤하게 Fuzzy c-partition U^0 를 생성하고 반복횟수 $t = 0$ 으로 한다. 멤버십 값 u_{ik} 이 주어졌을

때, 클러스터 중심 $v_i^{(t)}$ ($i=1, \dots, c$) 은 다음과 같이 계산되어 진다.

$$v_i^{(t)} = \frac{\sum_{k=1}^n (u_{ik}^{(t)})^m x_k}{\sum_{k=1}^n (u_{ik}^{(t)})^m} \quad (3)$$

새로운 클러스터 중심 $v_i^{(t)}$ 가 주어지면, 멤버쉽 값 $u_{ik}^{(t)}$ 은 아래와 같이 갱신된다.

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^c \left(\frac{\|x_k - v_j^{(t)}\|^2}{\|x_k - v_i^{(t)}\|^2} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (4)$$

이러한 반복 과정은 $|U^{(t+1)} - U^{(t)}| \leq \epsilon$ 을 만족하거나 사전에 정의된 반복 횟수에 도달하면 멈추게 된다.

3.2 Artificial Neural Network

Backpropagation 은 다수의 레이어를 갖는 ANN 을 학습시키기 위한 체계적인 방법이다.[3][4] Network 를 훈련시키는 목적은 입력 값과 출력 값으로 이루어진 훈련 쌍을 응용 시스템에 적용해 입력 값에 상응하는 출력 값을 얻기 위해 Network 의 가중치를 조절하기 위함이다. 이러한 훈련 과정이 수행되기 전에 먼저 초기화가 이루어져야 하는데 이러한 초기화를 위해서 모든 가중치는 불규칙한 작은 값으로 주어진다. 이러한 초기화는 Network 가 큰 가중치에 의해 포화되는 즉, Network 가 마비되는 것을 방지한다. Backpropagation 훈련 알고리즘은 아래와 같이 단계별로 구분되어진다.

1. 훈련 데이터로부터 입출력 쌍을 선택해 Network 의 입력 값을 받는다.
2. 현 입력 값에 대한 출력 값을 계산한다.
3. 입력 값에 대한 출력 값과 원래 입출력 쌍의 출력 값의 차(오차)를 계산한다.
4. Network 의 가중치를 3에서 구한 오차를 최소화 시킬 수 있도록 조절한다.
5. 각각의 훈련 쌍에 대하여 일정한 오차 수준을 만족할 때까지 단계 1 부터 단계 4 까지 반복 수행 한다.

IV. 학교 선별 시스템 모델링

학교 선별 시스템의 설계에 있어 가장 우선 고려되어야 할 사항은 TOEFL 점수이다. 이는 각

학교가 기본적으로 최소 TOEFL 점수를 제시하고 있기 때문이다.

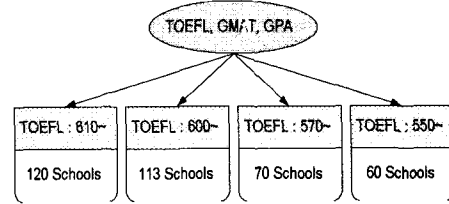


그림 2. TOEFL 점수에 근거한 그룹화

따라서 본 논문에서는 전체 120 개 학교를 TOEFL 점수에 따라 4 개의 그룹으로 분류하였다. 각 그룹은 그림 2 에 나타난 것과 같이 각각 120 개, 113 개, 70 개, 60 개의 학교로 구성하였다.

전체 학교를 TOEFL 점수에 따라 4 개의 그룹으로 분류한 다음, 각 그룹 내의 학교를 Fuzzy clustering 알고리즘을 적용하여 각각 15 개, 13 개, 10 개, 10 개의 소 그룹으로 다시 분류하였다. 이러한 재 분류 과정을 그림 3 에 나타내었다. 재 분류된 소 그룹의 그룹 번호와 각 학교의 GMAT, GPA 점수를 각각 출력, 입력 값으로 훈련 데이터를 생성해 Backpropagation 알고리즘으로 Network 를 훈련시켰다. 입력, 히든, 출력 뉴런으로 각각 2, 15, 4 개의 뉴런을 사용하였다. 각기 다른 4 개의 Network 를 최소 3,000 회의 epoch 과 0.001 의 제한 오차를 가지고 훈련시켰다.

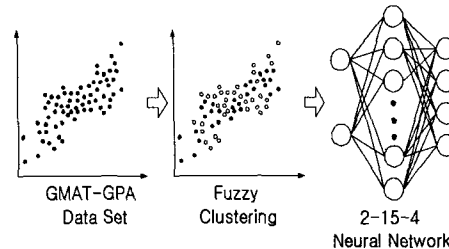


그림 3. Clustering 과 훈련 과정

위에서 설명한 학교 분류 시스템의 전체적인 시스템 구성과 입력 정보의 처리 과정을 그림 4 에 나타내었다. 학교 선별 시스템의 정보 처리 과정을 단계별로 나타내면 아래와 같다.

- Step 1. 미 경영대학원 입학 지원자의 TOEFL, GMAT, GPA 점수를 입력 받는다.
- Step 2. 입력 정보의 구간이 확인된다. (TOEFL: 550~670, GMAT: 450~740, GPA: 3.0~4.0)
- Step 3. TOEFL 점수에 따라 훈련된 4 개의 Network 중 하나가 선택된다.

Step 4. 입력 정보로부터 가장 적절한 그룹 번호가 선택된다.

Step 5. 그룹 번호에 따른 학교 명단이 주어진다.

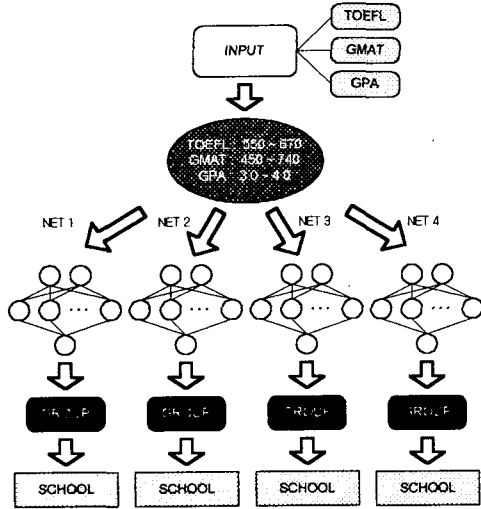


그림 4. 전체적인 시스템 구조

학교 선별 시스템의 출력 값은 그룹 번호와 이에 상응하는 학교 명단이 된다. 이때 출력된 그룹 번호는 우선 순위를 나타낸다. 다시 말해, 미 경영대학원 입학 지원자는 자신의 그룹 번호보다 작은 번호를 갖는 그룹에 포함된 학교에 대해서도 지원이 가능하며 입학 가능성은 높아지게 된다.

V. 성능 평가

본 논문에서 소개한 학교 선별 시스템의 경우, 절대적인 참 값이 존재하지 않기 때문에 그 성능을 구체적인 수치로 평가하기는 어렵다. 이러한 이유로 인해 본 시스템의 성능을 평가하기 위하여 종로에 위치한 사설 유학원을 방문해 유학 상담 경력 5년 이상의 현직 전문가들의 자문을 구했다. 몇 차례의 실제 학생들의 GPA, TOEFL, GMAT 점수를 입력해 출력 결과를 확인해 본 결과 매우 정확한 결과를 얻을 수 있었고 현업에 종사하는 전문가들도 긍정적인 의견을 보였으며 95% 이상 신뢰도를 보인다는 평가를 받았다. 또한 몇몇 전문가들로부터 학교 선정에 있어 중요한 요소는 GPA, TOEFL, GMAT 뿐 아니라 Essay와 직장 경력이기 때문에 이러한 부분에 대한 고려가 필요하다는 지적을 받았다. 그러나 전체적으로 본 논문에서 소개하는 시스템이 미 경영대학원 지원에 있어 지원자들에게 중요한 참고 자료가 될 수 있다는 평가를 받았다. 실제 본 학교 선별 시스템의 사용 예를 그림 5에 나타내었다.

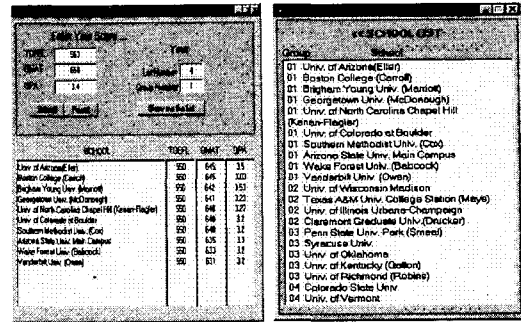


Figure 5. 학교 선별 시스템의 실 적용 예

VI. 결론

본 논문에서는 ANN을 미 경영대학원 지원자를 위한 학교 선별 시스템에 응용하였다. Fuzzy clustering을 통해 Network 훈련에 필요한 출력 값을 얻었고, 각 학교별 입학 요구 사항 중 객관적으로 수치화 할 수 있는 GPA, TOEFL, GMAT 성적을 입력 값으로 사용하였다. 그리고 이러한 ANN의 응용인 학교 선별 시스템을 전문가들의 시험을 통해 평가를 받았다.

비록 성능 평가는 긍정적이었으나 본 시스템에서 고려한 세가지 평가 요소만이 학교 선정에 있어 절대적이지 않다는 것은 의심의 여지가 없다. 그러나 Essay와 직장 경력은 객관적으로 평가되어지는 요소가 아닌 각 학교별 입학 사정 위원회에 의해 평가되어지는 주관적인 요소이다. 이러한 모호함으로 인해 본 시스템에서는 주관적인 평가 요소를 배제하였다.

본 논문에서 소개한 시스템의 출력 값으로 제시된 학교 명단에 대한 신뢰도가 절대적이지는 못하지만 산재해 있는 다양하고 불확실한 정보 속에서 난감해 하는 미국 소재 경영대학원 진학 희망자들에게는 매우 큰 도움이 되리라 생각된다.

참고문헌

- [1] E. Ruspini. A New Approach to Clustering. *Information and Control*. 15:22-32, 1969.
- [2] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [3] Widrow, B. and Stearns, S.D.. *Adaptive Signal Processing*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1985.
- [4] Rumelhart, D., et al., *Learning Representation by Back-Propagating Errors*. *Nature Col.*, 323, October, 1986