

입력자료 판별에 의한 데이터 마이닝의 성능개선

Performance Improvement of Data Mining by Input Data Discrimination

†이재식, ‡이진천

†아주대학교 경영대학 경영학부, ‡d2K Solutions

†(Tel) 0331-219-2719 (Fax) 0331-219-2190

†(E-mail) leejsk@madang.ajou.ac.kr

요 약

데이터 마이닝의 수행 예측 오차를 줄이기 위한 방법으로 하나의 문제를 여러 기법들을 결합하여 해결하고 있다. 본 연구에서는 새로운 결합모델을 제시하고 이를 통해 예측 오차를 감소시킬 수 있는 가능성을 제시한다. 제시된 결합모델의 성능을 검증하기 위해서 국내 자동차보험 회사의 고객데이터를 바탕으로 고객이탈 예측문제를 다루었다. 결합모델의 예측결과를 의사결정나무, 사례기반추론 그리고 인공지능망 중 하나의 기법만을 사용하여 예측한 결과와 비교 평가하였다. 평가 결과, 결합모델의 예측 적중률이 개별 기법의 예측 적중률보다 우수했다.

1. 서 론

정보 기술의 발달과 더불어 기업들은 많은 양의 데이터를 보유하게 되었다. 최근 이러한 데이터를 분석하여 경영활동에 활용하고자 하는 데이터 마이닝이 폭 넓게 적용되고 있다. 데이터 마이닝의 수행 시에 한가지의 데이터 마이닝 기법만을 사용하면 예측 오차를 감소시키는데 한계가 있기 때문에 여러 가지 다른 기법들을 결합하여 사용하기 시작했다. 본 연구의 목적은 새로운 결합모델의 방법을 제시하는 것이다. 본 연구에서 제시하는 결합모델은 세 개의 모델 즉, 판별모델, 기본모델 그리고 지원모델로 구성되어 있다. 판별모델은 의사결정나무에 의해 설계되었고, 기본모델은 사례기반추론 기법에 의해 설계되었으며 마지막으로 지원모델은 인공지능망 기법에 의해 설계되었다. 자동차 보험 가입

고객의 이탈예측 문제를 통해 본 연구에서 제시한 결합모델의 성능을 평가하였다. 합리적인 평가를 위해 결합모델 이외에 인공지능망, 의사결정나무, 사례기반추론의 기법을 각각 이용하여 세 개의 독립모델을 따로 설계하고 이들의 적중률과 결합모델의 적중률을 비교하였다.

2. 데이터 마이닝 모델의 결합 방법

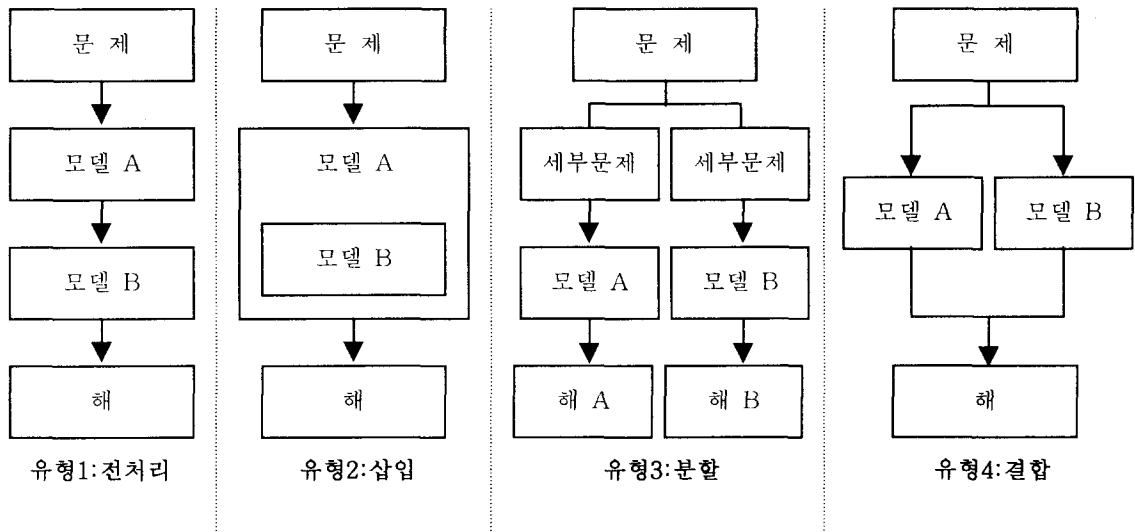
데이터 마이닝은 일반적으로 ‘아직 알려지지 않은(Unknown) 그렇지만 타당하고(Valid) 활용성 있는(Actionable) 정보(관계, 패턴, 규칙, 추세)를 데이터베이스로부터 찾아내는 과정’이라고 정의할 수 있다[Berry and Linoff, 1997]. 데이터 마이닝의 문제영역 및 응용분야는 매우 다양하다. 따라서 어떤 문제를 해결함에 있어서 단순히 하나의 기법만을 적용하여 문제를 해결하

는 것은 적합한 접근법이라고 할 수 없다. 즉, 여러 가지의 기법을 적용하여 그 적절성을 검사해야 됨은 물론이고 필요에 따라서는 여러 가지 기법을 결합해서 사용해야 하는 것이다.

하나의 문제를 해결하기 위하여 여러 가지 데이터 마이닝 기법들을 결합해서 사용하는 연구는 그 동안 다양하게 수행되어 왔다[이재식과 차봉근, 1999; Cho and Kim, 1995; Kuncheva et al., 1998; Kim and Shin, 1999]. 데이터 마이닝의 개별 기법들은 저마다의 단점들을 가지고 있다. 이러한 단점들을 보완하고 시스템의 정확도를 높이고자 하는 목적에서 모델의 결합이 시작된 것이다. 모델을 결합하는 방법에는 <그림 2-1>과 같이 다양한 방법들이 연구되었다.

모델결합 유형 1은 하나의 모델이 다른 모델의 효율적인 적용을 위해 전처리 기능을 담당하는 것이다. 예를 들면, 의사결정나무를 통해

속성을 선정한 후 인공신경망을 통해 예측하는 형태가 유형 1에 해당한다. 유형 2는 하나의 모델이 다른 모델의 내부에 포함되는 형태로 인공신경망 모델 구축 시 유전적 알고리즘을 활용하여 최적의 인공신경망 구조의 탐색을 도와주는 경우이다. 유형 3은 복잡한 문제영역에 대하여 해결 해야 될 문제를 세분화 한 후 각각 모델을 세워 문제를 해결하는 경우이다. 마지막으로 유형 4는 한 문제를 둘 이상의 모델로 각각 해결한 후 그 결과값들을 결합하여 편향성(bias)을 줄이고자 하는 방식이다[한인구와 신경식, 1999]. 본 연구에서는 세 가지의 데이터 마이닝 기법, 즉 인공신경망, 의사결정나무, 사례기반추론 기법을 결합하여 하나의 예측 모델을 구성하였다. 모델 결합의 유형은 유형 1과 유형 4가 합쳐져 있는 형태인데, 자세한 내용은 제 5절에서 기술한다.



<그림 2-1> 모델 결합의 4가지 유형

3. 고객이탈 예측을 위한 데이터

금융, 유통, 통신 등의 업종에서 고객이탈의 문

제가 심각한 기업 이슈로 대두되고 있다. 국내 보험산업은 정부의 보호하에서 높은 성장율을

보여 왔으나, 금융과 보험간의 규제완화, 보험 효율의 자율화 및 외국 보험업체의 국내 진출에 따라 이제 국내 보험 회사들은 고객확보 및 기존고객유지에 심혈을 기울이게 되었다. 본 연구에서는 이탈고객을 '보험가입 고객 중 보험만료일 시점으로부터 3개월 이내에 재가입을 하지 않은 고객'으로 정의한다.

본 연구에서 사용된 데이터는 1997년 8월부터 1998년 11월까지의 국내 A사 자동차 보험에 가입한 고객데이터이다. 이탈고객의 예측을 위해 먼저 A사의 운영계 데이터베이스로부터 이탈과 관련 있을 것으로 판단되는 53개의 속성으로 구성된 고객테이블을 생성하였다. 생성된 고객테이블은 약 110만개의 레코드를 포함하고 있으며, 이 중 이탈고객의 비율은 약 45.5%에 해당한다. 이 데이터에 대해서 잡음과 널 값 등을 적절한 형태로 변환 또는 제거 하는 데이터정제(Data Cleaning) 작업을 수행하였고, 또한 고객이탈 현상을 예측하는데 필요하다고 판단되는 새로운 파생속성(Derived Variable) 20여개를 생성하였다. 이렇게 모여진 70여개의 속성들로부터 본 연구에서는 해당 속성의 값에 따른 고객 이탈율의 차이 정도 및 자동차 보험 영역의 전문지식(Domain Knowledge)을 이용하여 속성을 선정하는 방법을 통하여 총 46개의 입력속성을 선정하였다. 선택된 속성들의 일부는 <표 3-1>과 같다.

<표 3-1> 고객이탈 예측을 위한 속성

구분	속성명	속성타입
1	고객 성별	범주형

4. 고객이탈 예측 모델의 설계

본 절에서는 고객이탈을 예측하기 위한 결합모

2	고객 연령	수치형
3	보험 가입 경력	수치형
4	전계약사	범주형
5	차량 연식	수치형
6	납입방법	범주형
7	수납형태	범주형
8	고객 직업 대분류	범주형
9	보험종목	범주형
10	차종약칭	범주형
11	보험기간내사고여부	범주형
12	대인담보가입여부	범주형
.	.	.

구축할 예측 시스템의 성능을 저하시키지 않는 범위 내에서 설계의 효율성을 높이기 위해서, 본 연구에서는 약 110만명의 고객 레코드 중 무작위 추출방법을 통해 30,000건의 고객데이터를 추출 하였는데, 추출한 30,000건의 데이터의 적합성 여부를 판정하여 부적합한 경우에는 다시 샘플링 작업을 수행하였다. 적합성 여부를 판정하는 기준은 '샘플링된 30,000건의 데이터 집합이 전체 고객 데이터 집합과 동일한 이탈비율을 가지고 있는가'이다. 본 연구에서 최종적으로 선택된 샘플 데이터 집합의 고객 이탈율은 45.6%로서 전체 고객 데이터 이탈율인 45.5%와 통계적으로 동일한 이탈 분포를 보인다고 말할 수 있다. 선택된 30,000건의 데이터는 다시 두 그룹으로 나누어 20,000건의 데이터는 모델의 학습을 위한 학습 데이터(Training Data)로 사용하였고 나머지 10,000건은 학습된 모델의 평가를 위한 테스트 데이터(Test Data)로 사용하였다.

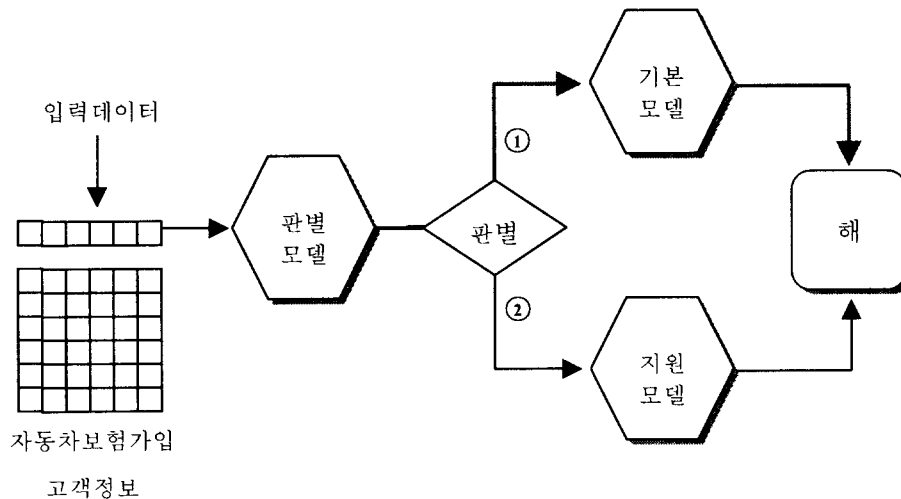
델을 설계하는 과정을 개괄적으로 설명하고, 상세한 설명은 제 5절에서 한다. 본 연구에서 제

시하는 결합모델은 <그림 4-1>과 같다.

먼저 입력데이터가 들어오면 판별모델을 통해 이 데이터가 기본모델과 지원모델 중 어떤 모델을 사용하여 예측하는 것이 더 적절한지를 판별하게 된다. 만약 판별모델을 통한 분석 결과 입력데이터가 기본모델에 더 적합하다고 판별되면 <그림 4-1>의 ①번 방향으로 예측 과정이 진행되어 기본모델을 사용하여 입력데이터에 대한 예측 결과를 얻게 된다. 마찬가지로 판별모델의 분석을 통하여 입력데이터가 지원모델에 더 적합하다고 판별되면, ②번 방향으로 예측과정이 진행되어 지원모델을 사용하여 입력데이터에 대한 예측 결과를 얻게 된다. 본 연구에서 제시하는 결합모델이 기존의 결합모델의 유형과 다른 점은 입력데이터의 특징 분석을 통해 예측을 수행할 모델을 선택하는 단계가 포함되어 있다는 점이다. 자동차 보험 가입 고객의 이

탈여부를 예측하기 위한 결합모델을 설계하기 위해서는 다음과 같은 총 4단계의 작업 과정이 필요하다.

- 제 1 단계 : 기본모델을 만든다. 결합모델을 구축하기 위한 처음 단계로서 인공지능 경망, 의사결정나무, 사례기반추론 기법을 통해 독립적인 고객이탈 예측모델을 설계한 후 가장 좋은 적중률을 보이는 모델을 기본모델로 선택한다.
- 제 2 단계 : 판별모델을 만든다. 제 1 단계에서 선택된 기본모델의 학습데이터에 대한 예측 적중여부를 새로운 목표속성으로 설정한 후, 각 입력데이터에 적합한 기본모델을 판별하는 모델을 구축하는 단계이다. 본 연구에서는 의사결정나무 알고리즘 중 하나인 C5.0 기법을 사용하여 판별모델을 설계한다.



<그림 4-1> 이탈고객 예측을 위한 결합모델의 구조

- 제 3 단계 : 지원모델을 만든다. 제 1 단계에서 선택된 기본모델이 잘못 예측한 데

이터만을 선별한 후 이 데이터를 가지고 새로운 예측모델을 설계한다. 이

단계에서 생성된 모델은 입력데이터가 기본모델에 적합하지 않다고 판별된 데이터에 대해서만 예측을 담당하는 지원모델로 사용된다. 본 연구에서는 지원모델의 설계를 위해 인공신경망 기법을 사용한다.

- 제 4 단계 : 결합모델을 완성한다. 제 1, 2, 3 단계에서 생성된 기본모델, 판별모델, 지원모델을 결합하여 최종적으로 자동차보험 가입 고객의 이탈여부를 예측하는 결합모델을 완성한다.

5. 독립모델과 결합모델의 설계 및 평가

5.1 독립모델의 설계 - 인공신경망

본 연구에서는 다계층퍼셉트론(Multi-Layer Perceptron)의 구조로 인공신경망 독립모델을 설계하였다[Nelson and Illingworth, 1991]. 입력층(Input Layer)에는 제 3절에서 선별된 46개의 속성을 사용하였고, 출력층(Output Layer)은 이탈여부를 판별하는 하나의 노드로 구성되었다. 은닉층(Hidden Layer)의 경우에 그 수의 설정에 이론적으로 제한이 없으나, 하나 혹은 2개의 층만으로도 모든 형태의 비선형 문제의 해결이 가능하므로 2개 이하의 범위 안에서 구조를 변형시키면서 실험을 하였다. 학습에는 백프로퍼게이션(Back-Propagation) 학습 알고리즘을 사용하였다.

모델의 실험은 은닉층의 개수, 은닉층 내의 처리요소의 개수들을 변화시키면서 학습율과 예

측율 변동추이를 관찰하였다(부록 <표 1>과 <표 2> 참조). 학습율이란 훈련용 데이터에 대한 적중률이고, 예측율이란 테스트용 데이터에 대한 적중률이다. 그 결과, 인공신경망을 이용한 자동차 보험 가입고객의 이탈 예측은 은닉층의 처리요소 개수가 30개인 단일 은닉층 구조에서 가장 좋은 예측율인 64.5%를 보여 주었다.

5.2 독립모델의 설계 - 의사결정나무

의사결정나무는 대상이 되는 집단을 몇 개의 소집단으로 구분하여 분류 및 예측을 수행하는 기법이다. 의사결정나무를 생성하는 과정은 분류 과정에서 주어지는 기준 즉, 나무의 깊이, 최종 노드의 개수 및 그 안의 사례의 개수, 가지의 분리방법, 가지치기 등에 의해 이루어진다. 의사결정나무는 가지의 분리방법에 따라 여러 종류의 알고리즘이 있는데, 본 연구에서는 이들 중 Quinlan에 의해 가장 최근에 개발된 C5.0 알고리즘을 사용하여 이탈예측 모델을 설계하였다. 가지의 분리기준으로는 이득률(Gain Ratio)을 사용하였다.

의사결정나무에서 생성된 규칙의 과잉적합(Overfitting)을 방지하기 위해서는 적절한 가지치기 작업 및 최종노드에 있어서의 적절한 사례 개수의 설정이 매우 중요하다. 그러므로 모델의 실험은 나무의 가지치기 엄격도의 값과 최종노드에 포함되는 최소 사례의 개수를 조정해가면서 각 모델이 보여준 학습율과 예측율을 비교 평가하면서 진행하였다(부록 <표 3> 참조).

C5.0 알고리즘을 이용한 의사결정나무 실험결과, 가지치기 엄격도 값이 85%이고 최종노

드의 최소 포함 사례 개수가 10개인 실험에서 가장 좋은 예측 성능을 보여주었다. 이때의 학습율과 예측율은 각각 79.8%와 63.2%였다.

5.3 독립모델의 설계 - 사례기반추론

사례기반추론은 새로운 문제가 주어지면 그 문제와 가장 유사한 문제를 과거 사례로부터 검색하여 그 사례의 결과값을 현재 문제의 결과값으로 그대로 또는 변형하여 활용하는 기법이다 [Aamodt and Plaza, 1996; Riesbeck and Schank, 1989]. 사례기반추론 모델을 구축하기 위해서는 인공신경망 및 의사결정나무 기법과는 달리 결과값이 알려진 과거의 사례들로 구성된 사례베이스(Case Base)의 구축이 선행되어야 한다. 본 연구에서는 1차 샘플링한 30,000건의 데이터 집합 이외에 추가적으로 30,000건을 더 추출하여 사례베이스를 구축하였다.

사례기반추론 기법에서는 새로운 입력사례와 사례베이스 안에 있는 사례와의 유사성을 측정하는 기준이 있어야 한다. 본 연구에서는 고객 이탈 문제영역의 도메인 지식(Domain Knowledge)을 이용하여 <표 5-1>과 같이 유사도 점수를 부여하였다.

적절한 사례기반추론 모델을 찾기 위해서 추천사례의 개수를 변화 시켜가면서 실험을 진행하였다. 추천사례란 입력된 새로운 사례에 대해 결과값을 결정해 주기 위해 사례베이스로부터 추출되는 과거 사례를 의미한다. 본 연구에서는 추천사례의 개수를 1, 5, 10, 15, 20, 25, 30개 순으로 변화 시켜가면서 실험을 진행하였다(부록 <표 4> 참조). 그 결과, 추천사례 개수를 10개로 했을 때 가장 좋은 성능을 보여 주었다. 이때의 학습율과 예측율을 각각 71.4%와 65.6%였다.

<표 5-1> 사례기반추론 모델의 유사도 점수 산정 기준 (일부)

번호	속성명	속성형	유사도 점수 부여 방법
1	고객 성별	범주형	If N = C Then S=10 Else S=0
2	고객 연령	수치형	If N-C <= 9 Then S= 10- N-C Else S = 0
3	보험 가입 경력	수치형	If N-C <= 9 Then S= 10- N-C Else S = 0
4	전계약사	범주형	If N = C Then S=10 Else S=0
5	차량 연식	수치형	If N-C <= 4 Then S= 10-(2* N-C) Else S = 0
.	.	.	.

(S = 유사도 점수, N = 현재 사례의 속성값, C = 과거 사례의 속성값)

5.4 기본모델의 설정

자동차 보험 가입 고객의 이탈 예측을 위한 결합모델의 완성을 위해 앞에서 구축한 3개의 예

측 모델, 즉 인공신경망, 의사결정나무, 사례기반추론 모델 중 가장 좋은 예측 성능을 보이는 모델을 기본모델로 선택하였다. <표 5-2>는 독립모델의 실험들에서 가장 좋은 예측 성능을 보여준 결과들이다.

<표 5-2> 독립모델의 이탈예측 성능

독립모델	학습율	예측율
인공신경망	69.8%	64.5%
의사결정나무	79.8%	63.2%
사례기반추론	71.4%	65.6%

<표 5-2>에서 알 수 있듯이 독립모델들 중에서 사례기반추론에 의한 예측 모델이 가장 좋은 예측율인 65.5%를 보여 주고있다. 따라서 사례기반추론 모델을 결합모델의 기본모델로 선정하였다.

5.5 판별모델의 설계

결합모델에 있어서 판별모델의 역할은 새로운 입력사례가 주어졌을 때 이 사례를 기본 모델로 예측하는 것이 적합한지 아닌지를 판별해주는 것이다. 본 연구에서는 다음의 과정을 통해 판별모델을 설계하였다.

- 과정 1 : 기본모델을 각 학습데이터에 적용하여 '예측적중여부'라는 새로운 속성을 생성한다. 실제 이탈여부가 사례기반추론에 의한 예측값과 같을 때에 '예측적중여부'에 'T' 값을, 다를 때에는 'F' 값을 부여한다.
- 과정 2 : 사례기반추론 모델의 학습을 위해

사용된 20,000의 데이터를 두 개의 데이터 집합으로 나눈다.

- 데이터 집합 1 : 판별모델의 훈련용 데이터 집합 (14,000건)
- 데이터 집합 2 : 판별모델의 테스트용 데이터 집합 (6,000건)

- 과정 3 : 예측적중여부 속성을 목표 속성으로 설정한 후 의사결정나무(C5.0)를 이용하여 예측 모델을 구축한다.

판별모델의 실험은 제 5.2절 의사결정나무 모델 설계에서 사용된 실험 방법과 동일한 방식으로 진행하였다. 그 결과, 가지치기 엄격도 값이 80%이고 최종노드에 포함되는 최소 사례 개수가 20개일 때 가장 좋은 성능인 학습율 87.1%, 예측율 79.8%를 보여주었다. 따라서 이때의 모델을 결합모델에 있어서의 판별모델로 선정하였다.

5.6 지원모델의 설계

지원모델은 판별모델에 의해 기본모델로 예측하는 것이 부적합하다고 판별된 입력사례에 한해서 예측을 수행하는 모델이다. 따라서 지원모델은 기본모델이 잘못 예측한 사례들만을 가지고 모델을 구축하였다. 모델 구축 과정은 다음과 같다.

- 과정 1 : 기본모델에 의해 잘못 예측된 사례들만을 추출하여 새로운 데이터 집합을 만든다.
- 과정 2 : 데이터 집합을 2개로 분리한다.
 - 데이터 집합 1 : 지원모델의 훈련용

데이터 집합 (4,004건)

- 데이터 집합 2 : 지원모델의 테스트용 데이터 집합 (1,716건)

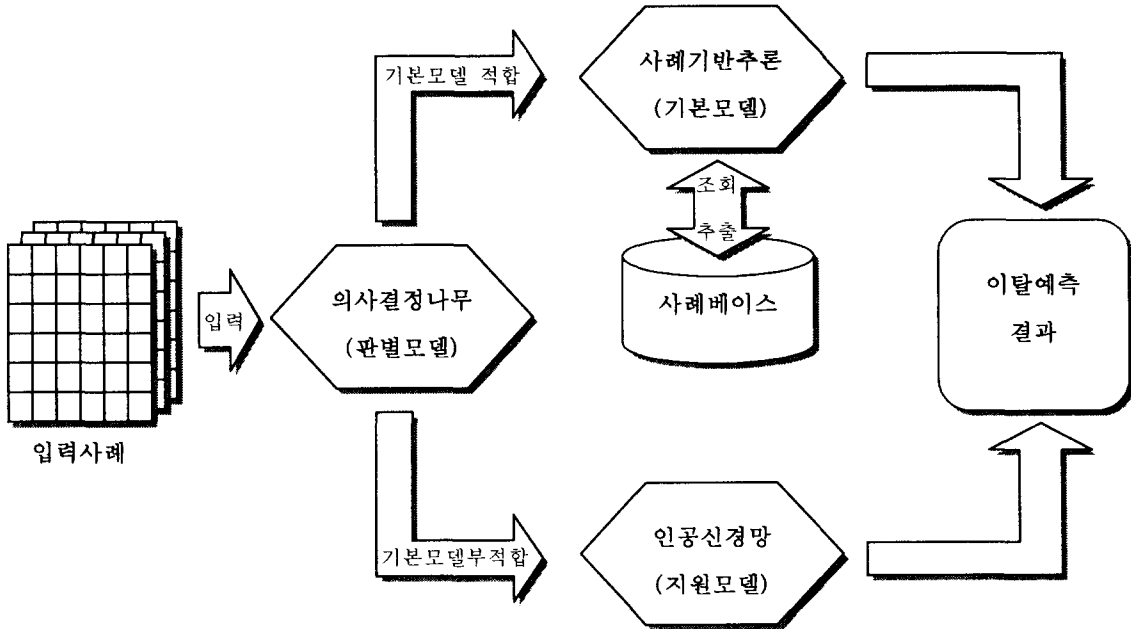
- 과정 3 : 이탈여부를 목표속성으로 설정한 후 인공지능망을 이용하여 지원모델을 구축한다.

요소의 개수를 변화시키면서 모델의 성능을 평가하였다. 그 결과, 은닉층의 처리요소 개수를 35개로 했을 때 가장 좋은 성능인 학습율 69.6%, 예측율 64.4%를 보여주었다. 따라서 이때의 모델을 결합모델에 있어서 지원모델로 선정하였다.

지원모델 설계를 위한 인공지능망은 제 5.1절에서 설계한 모델과 동일하게 다계층퍼셉트론과 백프로퍼게이션 알고리즘을 사용하였다. 은닉층의 경우 단일 은닉층 구조로 설정하고 처리

5.7 결합모델의 완성

자동차보험 가입고객의 이탈예측을 위한 결합모델의 구조는 <그림 5-1>과 같다.



<그림 5-1> 자동차보험 가입고객 이탈예측을 위한 결합모델의 구조

결합모델의 평가에서는 독립모델의 평가에 사용된 10,000건의 데이터를 동일하게 사용하

였다. 결합모델의 예측 결과와 독립모델의 예측 결과와의 비교는 <표 5-3>과 같다.

테스트용 데이터에 대한 결합모델의 예측율은 71.8%로서 독립모델에서 가장 좋은 예측결과를 보여준 사례기반추론의 예측율 65.6% 보다 약 6% 포인트가 높은 수치를 보여주었다. 이러한 결과는 사례기반추론 모델만을 사용할 경우 놓칠 수 있는 데이터의 숨은 패턴을 지원 모델인 인공지능망이 예측한 결과라고 생각할 수 있다.

6. 결 론

데이터 마이닝의 수행 시에 단일모델만을 적용하면 예측 오차를 감소시키는데 한계가 있기 때문에 결합모델이 사용되기 시작했다. 본 연구에서는 데이터 마이닝 시스템 설계 시 그 성능을 개선시킬 수 있는 새로운 형태의 결합모델을 제시하였다. 자동차보험 가입 고객의 이탈예측 문제를 통해 본 연구에서 제시한 결합모델의 성능을 인공지능망, 의사결정나무, 사례기반추론의 기법을 각각 독립적으로 사용했을 때의 성능과 비교 평가하였다.

평가 결과 독립모델들의 예측율은 사례기반

추론 모델이 65.6%로 가장 높았으며, 인공지능망 모델이 64.5%, 의사결정나무 모델이 63.2%이었다. 본 연구에서 제시한 결합모델로 이 독립모델들을 결합한 결과 예측율은 71.8%로서 사례기반추론 독립모델보다 약 6% 포인트 증가하였다. 물론 본 연구에서 제시된 모델의 결합방법이 일반적인 모든 문제 영역에 대하여 적합하지는 않을 것이다. 그러나 본 연구에서 제시된 결합모델로써 해의 패턴이 복잡한 특정 문제영역에서 예측률을 증가시킬 수 있다는 가능성을 제시하였다는 점에서 의의가 있다고 할 수 있겠다.

본 연구의 한계점 및 향후 연구과제를 지적하면 다음과 같다. 결합모델의 구성할 때에 다른 기계학습 기법 및 통계적인 방법 등 다양한 기법들을 사용하여 각 구성요소들을 구축한 후에 가장 성능이 좋은 조합으로 결합모델을 구축하는 연구가 필요하다. 또한, 결합모델의 평가를 위해 세 개의 독립모델을 설계하였는데 모두가 기계학습에서 사용되는 기법들이었다. 좀 더 타당한 평가를 위해서는 통계적인 기법의 모델들과도 비교 평가했어야 했다.

<표 5-3> 고객이탈 예측 모델별 예측 결과

구분		학습		테스트	
		데이터(건)	학습율(%)	데이터(건)	예측율(%)
독립 모델	인공지능망	20,000	69.8%	10,000	64.5%
	의사결정나무	20,000	79.8%	10,000	63.2%
	사례기반추론	20,000	71.4%	10,000	65.6%
결합모델		해당 없음		10,000	71.8%

참 고 문 헌

- [1] 이재식, 차봉근, “유전적 알고리즘을 이용한 인공신경망의 구조 설계,” 한국경영과학회지, 제24권 제3호, pp.49-62, 1999.
- [2] 한인구, 신경식, “지능형 중소기업 신용평가시스템의 개발 및 활용:보람은행의 사례를 중심으로,” 한국경영정보학회 발표논문집, pp. 51-61, 1999.
- [3] Aamodt, A. and E. Plaza, “Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches,” Artificial Intelligence Communications, Vol. 7, No. 1., 1996.
- [4] Berry, M. and G. Linoff, Data Mining Techniques for Marketing, Sales, and Customer Support, John Wiley & Sons, Inc., 1997.
- [5] Cho, S. B. and S. H. Kim, “Multiple Network Fusion using Fuzzy Logic,” IEEE Trans. Neural Networks, Vol.6, pp. 497-501, 1995.
- [6] Kim, S. H. and S. W. Shin, “Hybrid Learning Architecture for Advanced Data Mining: An Application to Binary Classification for Fraud Management,” Journal of Information Technology Application, vol. 1, pp.173-221, 1999.
- [7] Kuncheva, L., I. C. Bezdek and M. A. Shutton, “On Combining Multiple Classifiers by Fuzzy Templates”, International Conference on Artificial Neural Network, IEEE, pp.193-197, 1998.
- [8] Nelson, M. M. and W. T. Illingworth, A Practical Guide to Neural Nets, Addison-Wesley Publishing, Inc., 1991.
- [9] Riesbeck, C. K. and R. L. Schank, Inside Case-Based Reasoning, Hillsdale, N, J. Lawrence Erlbaum Associates, 1989.

부 록

<표 1> 한 개의 은닉층 구조에서
처리요소 개수 변화에 따른
학습율 및 예측율

실험 번호	처리요소 개수	학습율 (%)	예측율 (%)
1	10	60.7	58.5
2	15	64.5	61.2
3	20	64.5	60.3
4	25	68.5	62.1
5	30	69.8	64.5
6	35	68.8	62.1
7	40	65.6	60.3
8	45	64.5	61.7
9	50	63.3	60.3

<표 2> 두 개의 은닉층 구조에서
처리요소 개수 변화에 따른
학습율 및 예측율

실험 번호	처리요소 개수		학습율 (%)	예측율 (%)
	은닉층1	은닉층2		
1	10	10	64.5	60.7
2	10	15	60.3	57.8
3	10	20	63.4	59.8
4	10	25	63.9	61.2
5	10	30	65.4	61.9
6	15	10	63.8	56.8
7	15	15	62.1	58.2
8	15	20	64.9	60.7
9	15	25	63.3	53.8
10	15	30	60.8	57.7
11	20	10	67.5	64.2
12	20	15	66.3	63.8
13	20	20	68.3	61.9
14	20	25	67.7	63.7
15	20	30	63.8	60.7
16	25	10	62.8	59
17	25	15	64.5	61.4
18	25	20	64.8	60.5
19	25	25	63.6	60.2
20	25	30	64.7	61.6
21	20	10	66.0	63.1
22	20	15	65.7	63.3
23	20	20	65.3	59.2
24	20	25	67.9	63.8
25	20	30	63.6	61.0

<표 3> 의사결정나무(C5.0) 모델의
학습율 및 예측율

실험 번호	가지치기 엄격도	최소포함 사례개수	학습율 (%)	예측율 (%)
1	75	10	64.4	57.6
2	75	20	67.1	59.7
3	75	30	66.4	57.3
4	75	40	72.9	62.3
5	80	10	65.1	58.8
6	80	20	78.7	62.8
7	80	30	75.3	60.6
8	80	40	75.0	60.8
9	85	10	79.8	63.2
10	85	20	80.1	61.1
11	85	30	77.1	62.7
12	85	40	75.6	61.0
13	90	10	73.8	60.4
14	90	20	76.7	61.2
15	90	30	73.8	59.8
16	90	40	75.9	60.1
17	95	10	80.2	57.0
18	95	20	82.3	58.1
19	95	30	79.2	60.1
20	95	40	78.5	58.9

<표 4> 추천사례 개수 변화에 따른 사례기반추론 모델의 학습율 및 예측율

추천사례수	1	5	10	15	20	25	30
학습율(%)	68.8	69.3	71.4	72.1	68.5	69.6	70.5
예측율(%)	60.9	62.3	65.6	64.3	62.7	63.0	64.5