

## 의사결정트리를 사용한 방학 중의 효과적 학습 방법\*

김혜숙<sup>○</sup> 문양세 김진호  
강원대학교 컴퓨터과학과

{hskim, ysmoon, jhkim}@kangwon.ac.kr

### Decision Tree-based Analysis to Effective Studying Methods during Vacation

Hea-Suk Kim<sup>○</sup>, Yang-Sae Moon, and Jinho Kim

Dept. of Computer Science, Kangwon National University

#### 요 약

본 논문에서는 방학 중 효과적인 학습 방법을 분석하기 위해 의사결정트리를 사용한 접근법을 제안한다. 이를 위해, 우선 학업성취도에 영향을 미치는 방학 중의 학습방법에 대한 다양한 요소를 도출한다. 다음으로, 의사결정트리를 사용하기 위한 데이터 변환 및 분석 방법을 제안한다. 마지막으로, 설문조사를 통해 수집한 현실의 구체적 데이터에서 의사결정트리를 생성한다. 중학교 학생들에 대한 설문조사를 분석한 결과, 세 가지 의미 있는 결과를 도출하였다. 첫째, 인터넷 학습사이트 이용은 성적 하락에 영향을 미치는 것으로 나타났다. 둘째, 성적 변화에 큰 영향을 미칠 것으로 예상했던 과외는 실제로 큰 영향을 미치지 못하는 것으로 나타났다. 셋째, 다양한 학습방법의 병행은 오히려 성적 하락의 요인이 되는 것으로 파악되었다. 본 논문에서 제시한 분석 방법 및 결과는 학생들의 방학 중 생활 지도나 학습 계획 수립에 많은 도움이 될 수 있다고 사료된다.

#### 1. 서 론

데이터 마이닝은 대용량 데이터에서 정보 및 지식을 발견하는 대표적 분석 기법 중 하나이다[1]. 정보 기술의 발전에 따라 기업들은 수십 GB 이상의 방대한 데이터를 관리하기에 이르렀고, 기업의 관심은 데이터를 관리하는 개념을 넘어 데이터에서 보다 가치있는 정보나 지식을 찾고자 하는 부분으로 집중되고 있다. 이러한 요구에 따라 1990년대 초반부터 데이터 마이닝 연구가 활발하게 진행되었다[1]. 그 결과, SAS Enterprise Miner[2], IBM Intelligent Miner[3], DBMiner[4] 등의 데이터 마이닝 제품이 상용화되었다.

본 논문에서는 교육 분야의 학업성취도 분석에 데이터 마이닝 기법을 활용한다. 교육에서의 학업성취도(academic achievement)란 학습의 결과로 지식과 기능을 습득하는 과정이나 결과를 의미한다[5]. 즉, 학업성취도는 학습에 의해 얻어진 교과 성적이나 학교 교육활동을 통해 얻어지는 모든 교육성과이다. 이러한 학업성취도는 교육과정을 평가하고 다음 과정을 제대로 이수할 수 있는지를 판단하는 자료로 사용된다. 학업성취도를 측정하는 교육평가는 학생 개개인이 성취해야 할 교육 목표의 성취 정도를 점검하고 그 결과를 평가자와 피평가자에게 제공해준다. 그리고 이를 통하여 평가자와 피평가자의 교육적인 노력 및 의사결정을 도와주는 역할을 한다[5].

본 연구는 방학 중의 학습방법과 학업성취도의 관계 분석에 초점을 둔다. 현재 우리나라에서는 학업성취도 상승을 위해 학교 수업 이외에 과외, 학원수강, 교육방송 등의 다양한 교육 방법들이 동원되고 있다. 이러한 현실에서, 교육평가의 기능이 제대로 발휘되기 위해서는 다양한 교육 방법만큼 다양한 관점에서의 학업성취도 분석이 이루어져야 한다. 그런데, 최근까지 진행되어 온 학업성취도 연구는 주로 정규학기 중 학습방법이나 활동에 집중되어 왔다. 특히, 1년 중 4개월에 해당하는 방학과 학업성취도의 관계에 대한 연구는 미비한 형편이다. 더구나, 최근 방학은 학습의 연장선이라는 관점에서 교사들이나 학부모들의 높은 관심을 받고 있다. 이에 따라 본 연구에서는 방학 중의 학습방법이 학업 성취도에 미치는 영향을 체계적으로 분석한다.

학생들의 방학 중 효과적인 학습 방법을 분석하기 위하여 본 논문에서는 데이터 마이닝 기법 중 의사결정트리(decision tree)

를 적용한다. 이를 위해, 다음과 같은 네 단계로 연구 및 분석을 수행한다. 첫째, 데이터 마이닝 기법을 적용하기 위해 학업성취도에 변화를 줄 수 있는 항목을 도출하고, 이들 항목을 기반으로 데이터 분석 방법을 구체화한다. 둘째, 방학 중 학습방법이 학업성취도에 영향을 미치는 요소를 도출하기 위해 1) 과거성적, 2) 방학 중 학습방법, 3) 현재성적으로 구성되는 구체적 데이터를 설문조사를 통해 수집한다. 셋째, 수집한 데이터를 대상으로, 데이터 변환 및 정제를 거쳐, 상용 데이터 마이닝 도구인 SAS사의 Enterprise Miner(간략히 E-Miner라 한다)를 사용하여 의미있는 규칙과 패턴을 추출하고 해석한다.

#### 2. 관련 연구

학생들의 학업성취도에 관한 여러 연구가 교육 분야에서 진행되어 왔다[8-10]. 참고문헌 [8]에서는 수학 과목에서 교수 학습 태도에 따른 학업성취도 변화를 분석하였다. 이 연구에서는 설문조사 분석을 통해, 수학 과목의 성취도가 높은 학생은 수학에 대한 학습 습관과 학습 동기가 뚜렷하고 끈기를 가지고 열심히 학습한다는 결과를 도출하였다. 그리고 참고문헌 [9]에서는 MBTI(Myers-Briggs Type Indicator) 성격 유형론을 토대로 학생의 성격 유형에 따른 학업성취도를 분석하였다. 이 연구에서는 성격의 기본 유형과 학업성취도 사이에는 밀접한 관계가 있다는 결과를 도출하였다. 마지막으로, 참고문헌 [10]에서는 ICT(Information Communication Technology)를 활용한 학습에서의 학습자 특성에 따른 학업성취도를 분석하였다. 이 연구에서는 ICT 교수-학습 환경에서 정서적 요인, 사회적 요인 등이 학업성취도에 미치는 영향을 분석하였다.

그러나, 지금까지의 학업성취도 관련 연구들은 주로 정규 학기의 학업성취도 분석에 연구의 초점이 맞추어져 있었다. 즉, 본 연구에서 다루는 방학 기간 중의 학업성취도 변화 분석은 현재까지 연구된 바가 없다. 경험적으로 볼 때, 방학 중 학습방법이나 생활습관은 학업성취도 향상에 큰 영향을 미친다. 따라서, 본 논문은 방학 전후의 학업성취도 변화 요인을 체계적으로 분석한다는 점에서 기존 연구와 차별성을 가진다.

교육 분야에 데이터 마이닝을 적용한 참고문헌 [6]의 연구에서는 고등학생을 대상으로 1학기 성적과 학원수강 여부, 수업시간의 학습태도를 이용해 2학기 성적을 예측하는 의사결정트리를 생성하였다. 그 결과, "상위권 학생들의 경우, 다음 학기

\* 2005년도 강원대학교 학술연구조성비로 연구하였음.

성적에 수업태도나 학원수강이 영향을 미치지 않지만, 중하위권 학생들의 경우는 성적에 영향을 미친다."는 지식을 찾아내기도 하였다. 그러나, 이 연구에서는 학원수강 여부와 수업시간의 학습태도라는 두 가지 단편적 정보만 사용했기 때문에, 학업성취도에 영향을 줄 수 있는 과외나 인터넷 학습 등의 여러 요소들이 고려되지 않은 문제점이 있다. 또한, 학기와 학기 사이의 방학 중 학습방법이 전혀 고려되지 않은 문제점이 있다. 반면에, 본 논문에서는 학원수강 이외에 과외, 교육 방송 시청 등 여러 요소들을 고려하고, 정규 학기가 아닌 방학 기간을 학업성취도가 변화하는 주요 기간으로 삼는다.

3. 방학 중 학습방법과 학업성취도의 관계 분석

3.1 마이닝 항목 도출

방학 중 학습방법이 학업성취도 변화에 미치는 영향을 분석하기 위하여 본 절에서는 학업성취도에 영향을 미칠 수 있는 요소들을 도출한다. 먼저, 방학 전후의 학업성취도 변화를 판단하기 위한 요소로 학생 개개인의 방학 전 성적과 개학 후 성적을 사용한다. 그리고 과외, 학원수강 등과 같이 학업성취도에 영향을 줄 가능성이 높은 요소들을 학업성취도관련항목으로 선정한다.

학업성취도 변화를 측정하는 요소로는 **상승, 유지, 하락**을 사용한다. 즉, 방학 전 성적과 개학 후 성적을 비교하여 성적이 높아진 학생을 상승으로 분류한다. 예를 들어, 방학 전 성적이 '미'였던 학생이 개학 후 '우'로 높아졌으면, 그 학생은 방학을 전후로 학업성취도가 상승한 것으로 본다. 같은 방법으로, 방학 전 성적과 개학 후 성적을 비교하여, 성적이 동일한 학생을 유지로, 성적이 낮아진 학생을 하락으로 설정한다.

학업성취도관련항목은 직관적으로 생각했을 때 학업성취도에 영향을 줄 가능성이 높은 요소들로 선정한다. 즉, 학생들이 학업성취도 상승을 위하여 의도적으로 실시하는 학습방법을 대상으로 한다. 학업성취도관련항목에는 학원수강, 집단 또는 개인 대상의 사설과외, 교육방송 시청, 학습지에 의한 학습, 인터넷 학습사이트 이용 등이 있다. 이들 학업성취도관련항목들은 방학 전후의 성적과 함께 국어, 영어, 수학 등 각 과목별로 데이터를 수집한다.

3.2 의사결정트리

본 논문에서는 방학 중 효과적인 학습방법을 분석하기 위해서 분류(classfication)의 대표적 기법인 의사결정트리를 사용한다. 의사결정트리는 수집된 과거의 데이터 레코드들을 분석하여 이들 사이에 존재하는 패턴이나 분류 모델을 트리 구조로 도표화한 후, 이를 사용하여 분류와 예측을 수행하는 방법이다[1,6,7].

의사결정트리는 방학 중 어떠한 학습방법들이 학업성취도 변화에 영향을 미치는지 파악하기 위해 사용한다. 즉, "방학 중 어떤 학습방법을 사용하면 성적이 향상되는가?"와 같은 형태의 분석을 위해 의사결정트리를 사용한다. 본 연구에서는 제3.1절에서 도출된 요소들을 대상으로 의사결정트리를 생성한다. 의사결정트리의 목표변수는 방학 전후의 성적 변화, 즉, 상승, 유지, 하락으로 두고 트리를 생성한다.

본 연구에서는 의미있는 의사결정트리를 생성하기 위하여 E-Miner가 생성한 트리에 두 가지 전지 규칙(pruning rule)을 적용한다. 다음 두 가지의 전지 규칙 사용 목적은 E-Miner에 의해 생성된 트리에서 의미없는 노드들을 제거하기 위해서이다.

**전지 규칙 1:** 단말 노드에 속하는 개체 수(학생 수)는 전체 입력 데이터 수의 5% 이상이어야 한다. □

전지 규칙 1을 적용하는 이유는 단말 노드의 개체 수가 5% 미만으로 너무 작은 경우, 전체 학생에서 해당 학생들이 차지하는 비율이 너무 작아 규칙으로서 의미가 없기 때문이다. 즉, 5% 미만에 해당하는 일부 학생에게만 적용되는 규칙은 일반적인 규칙이라 보지 않는다.

**전지 규칙 2:** 루트 노드와 단말 노드의 성적 변화율 차이는 5%

이상이어야 한다. □  
전지 규칙 2를 적용하는 이유는 루트에서 단말까지의 성적 변화율 차이가 5% 미만으로 너무 작은 경우, 성적의 변화가 확실하다고 판단하기 어렵기 때문이다. 예를 들어, 성적이 '수'였던 학생의 5% 미만이 개학 후 성적이 '우'로 바뀐 것을 의미있는 규칙으로 보지 않는다.

4. 실험 및 분석

4.1 데이터 수집

학생들의 방학 중 효과적인 학습방법을 분석하기 위해 설문조사를 통해 현실의 구체적 데이터를 수집하였다. 설문조사서는 제3.1절에서 도출한 항목들을 기반으로 작성하였다.

설문조사는 강원도 춘천지역 중등학교 2학년 및 3학년 학생을 대상으로 하였다. 강원도 춘천지역은 2005년 현재 고교 비평준화 지역으로 평준화와 관련이 적은 중학교 학생을 설문 대상으로 하였다. 모집단의 학생 수는 10,790명이었으며 신뢰도 90%, 표본오차범위 5%인 경우의 표본 집단수는 371명이었다. 본 논문의 설문조사에는 500여명 이상의 학생들이 참여하였으므로, 신뢰도 90% 및 표본오차범위 5% 수준을 만족한다.

설문조사는 온라인 설문조사 전문업체인 리서치중앙[23]에 설문을 의뢰하였으며, 사이트는 여자중학교를 중심으로 홍보하였다. 데이터 수집을 위한 설문조사 기간은 2005년 5월 17일부터 2005년 5월 31일까지의 15일로 하였다. 설문에 응답한 학생은 총 536명이었는데, 일부 항목에만 응답한 불완전 데이터와 상대적으로 비율이 적은 남학생 데이터를 제외한 여학생 462명의 데이터를 사용하였다.

4.2 실험 결과

제3.2절에서 설명한 바와 같이, 트리는 학원수강, 과외, 교육방송 시청, 학습지 학습, 인터넷 학습 사이트 이용 등의 학업성취도관련항목을 입력 변수로 생성한 의사결정트리이다.

국어 과목에 대한 전체 트리 생성 결과는 그림 3과 같다. 그림을 보면, E-Miner에 의해 생성된 트리에서 네 개의 단말 노드가 전지 규칙 1과 전지 규칙 2에 의해 삭제되었음을 알 수 있다. 결국, 그림 3에서는 두 개의 의미있는 단말 노드가 생성되고, 이들 단말 노드를 논리곱(logical AND)형태의 지식으로 나타내면 표 1과 같다. 표의 지식을 해석하면, "교육방송은 국어 성적의 상승을 인터넷 학습 사이트는 성적을 하락시킨다."고 이해할 수 있다.

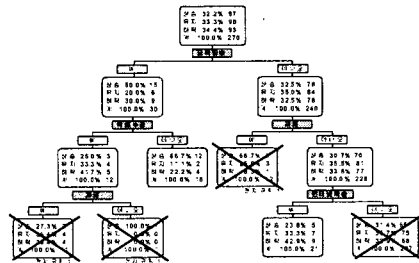


그림 3. 국어 과목의 전체 트리.

표 1. 국어 의사결정트리에 대한 논리곱 형태의 지식.

분류	조	건
상승	(교육방송=예) ∧	(학원수강=아니오)
하락	(교육방송=아니오) ∧	(과외=아니오) ∧ (인터넷학습=예)

다음으로, 그림 4는 영어 과목에 대해 생성한 트리를 나타낸다. 국어 과목과 마찬가지로 트리를 생성한 후 두 가지 전지 규칙을 사용하여 의미 없는 단말 노드를 제거하였다. 그리고 표 2는 트리의 단말 노드들을 대상으로 논리곱 형태의 지식을

표현한 것이다. 표를 해석하면, “과외를 하면 영어 성적이 유지되며, 학원수강, 과외, 인터넷 학습 사이트의 세 가지를 모두 하지 않으면 성적이 오히려 상승한다.”는 지식으로 이해할 수 있다.

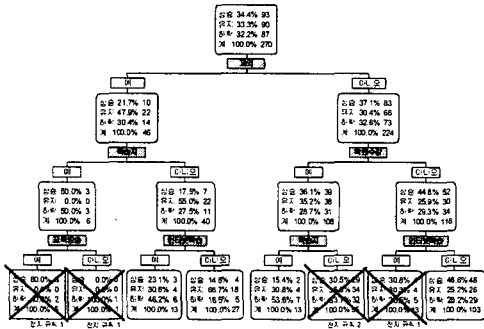


그림 4. 영어 과목의 전체 트리.

표 2. 영어 의사결정트리에 대한 논리곱 형태의 지식.

분류	조 건
상승	(과외=아니오) ∧ (학원수강=아니오) ∧ (인터넷학습=아니오)
유지	(과외=예) ∧ (학습지=아니오) ∧ (인터넷학습=아니오)
하락	(과외=예) ∧ (학습지=아니오) ∧ (인터넷학습=예)
	(과외=아니오) ∧ (학원수강=예) ∧ (학습지=예)

마지막으로, 그림 5는 수학 과목에 대해 생성한 트리를 나타낸다. 마찬가지로, 전치 규칙 1 및 2를 사용하여 의미 없는 단말 노드를 제거하였다. 표 3은 그림의 단말 노드를 대상으로 논리곱 형태의 지식을 추출한 것이다. 표의 지식을 해석하면, “학원수강을 하면 수학성적이 유지된다.”고 볼 수 있다.

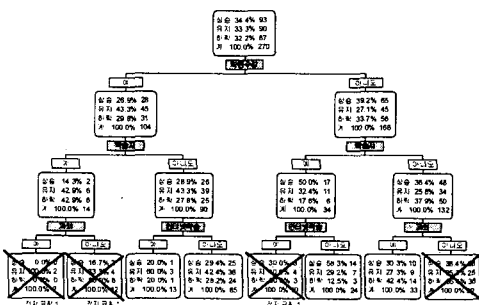


그림 5. 수학 과목의 전체 트리.

표 3. 수학 의사결정트리에 대한 논리곱 형태의 지식.

분류	조 건
상승	(학원수강=아니오) ∧ (학습지=예) ∧ (인터넷학습=아니오)
유지	(학원수강=예) ∧ (학습지=아니오) ∧ (인터넷학습=예)
하락	(학원수강=예) ∧ (학습지=아니오) ∧ (인터넷학습=아니오)
	(학원수강=아니오) ∧ (학습지=아니오) ∧ (과외=예)

지금까지의 살펴 본 국어, 영어, 수학 세 과목의 전체 트리에서 추출된 지식을 종합하면 다음과 같은 결과를 얻을 수 있다. 첫째, 인터넷 학습 사이트 이용은 성적 하락의 원인으로 작용하고 있다. 둘째, 성적 향상에 큰 영향을 줄 것으로 예상한 과외가 실제로는 큰 영향을 미치지 못한다. 셋째, 다양한 학습 방법(교육 방송 시청, 학원수강, 과외 등)을 병행할 경우 오히려 성적이 하락한다.

## 5. 결 론

본 연구에서는 방학 중 효과적인 학습방법을 데이터 마이닝 기법을 사용하여 분석하였다. 현재까지 진행된 학업성취도 관련 연구에서는 학생의 학업성취도 향상을 위한 방학 중 학습방법에 관한 구체적인 접근법이 없었다. 본 연구에서는 데이터 마이닝 기법을 사용하여 다음과 같은 구체적 분석 방법을 제안하였다. 우선, 학업성취도에 영향을 미치는 방학 중 학습방법에 대한 다양한 요소들을 도출하였다. 그리고 이들 요소와 학업성취도 변화의 관계 분석을 위해서 데이터 마이닝 기법들 중 의사결정트리를 사용하였다. 또한, 현실의 구체적 데이터 수집을 위해 중학교 학생들을 대상으로 방학 전후 성적과 방학 중 활동에 대한 설문조사를 수행하였다. 다음으로, 의사결정트리를 생성하여 방학 중 효과적인 학습방법을 분석하였다.

방학 중 효과적인 학습방법 분석을 위해 생성한 의사결정트리에서 얻은 지식을 요약하면 다음과 같다. 첫째, 인터넷 학습사이트 이용은 일반적으로 성적을 하락시키는 것으로 나타났다. 둘째, 성적 변화에 큰 영향을 미칠 것으로 예상했던 과외는 실제로 큰 영향을 미치지 못하는 것으로 나타났다. 셋째, 다양한 학습방법(과외, 학원수강, 교육 방송 시청)을 병행할 경우 오히려 성적이 하락하는 것으로 나타났다.

본 논문에서는 설문조사와 데이터 마이닝 기법을 사용하여 방학 중 효과적인 학습방법을 분석하였다. 이와 같은 연구 결과는 추후 학생들의 방학 중 학습 계획 수립이나 방학 중 학습 지도에 매우 우수한 모형을 제시한다고 사료된다.

## 참고 문헌

- [1] Chen, M.-S., Han, J., and Yu, P. S., "Data Mining: An Overview from a Database Perspective," *IEEE TKDE*, Vol. 8, No. 6, pp. 866-883, Dec. 1996.
- [2] <http://www.sas.com/korea>.
- [3] <http://www-306.ibm.com/software/data/iminer>.
- [4] <http://www.dbminer.com>.
- [5] 김신영, 백중근, 채선희, 국가교육과정에서 근거한 평가 기준 및 도구 개발, 한국교육과정평가원, 1998.
- [6] 배재호, 데이터 마이닝을 이용한 학업성취도 분석, 경희대학교 교육대학원 석사학위논문, 2001.
- [7] Agrawal, R., et al., "An Interval Classifier for Database Mining Applications," In *Proc. the 18th Int'l Conf. on VLDB*, Vancouver, Canada, pp. 560-573, Aug. 1992.
- [8] 김동화, 조영재, "수학 교수, 학습 태도에 따른 학업성취도 수준에 관한 분석," *교육이론과 실천*, 제14권, 제1호, pp. 115-132, 2004.
- [9] 김병식, 김민정, "청소년의 성격유형이 학업성취도에 미치는 영향에 관한 연구," *한국인간관계학보*, 제7권, 제1호, pp. 1-20, 2002.
- [10] 김정경, "정보통신기술(ICT)을 활용한 교수-학습에서 학습자 특성이 학업성취도에 미치는 영향," *한국컴퓨터교육학회 논문지*, 제7권, 제2호, pp. 47-56, 2004.
- [11] 장남식, 홍성환, 장재호, 데이터 마이닝, 대청, 1999.
- [12] <http://www.research.joongang.com>.