J. Soc. Korea Ind. Syst. Eng Vol. 38, No. 4 : 45-55, December 2015
http://dx.doi.org/10.11627/jkise.2015.38.4.45
ISSN : 2005-0461(print)
ISSN : 2287-7975(online)

Curriculum Mining Analysis Using Clustering-Based Process Mining

Woo-Min Joo · Jin Young Choi†

Department of Industrial Engineering, Ajou University

군집화 기반 프로세스 마이닝을 이용한 커리큘럼 마이닝 분석

주우민 · 최진영†

아주대학교 산업공학과

In this paper, we consider curriculum mining as an application of process mining in the domain of education. The basic objective of the curriculum mining is to construct a registration pattern model by using logs of registration data. However, subject registration patterns of students are very unstructured and complicated, called a spaghetti model, because it has a lot of different cases and high diversity of behaviors. In general, it is typically difficult to develop and analyze registration patterns. In the literature, there was an effort to handle this issue by using clustering based on the features of students and behaviors. However, it is not easy to obtain them in general since they are private and qualitative. Therefore, in this paper, we propose a new framework of curriculum mining applying K-means clustering based on subject attributes to solve the problems caused by unstructured process model obtained. Specifically, we divide subject's attribute data into two parts : categorical and numerical data. Categorical attribute has subject name, class classification, and research field, while numerical attribute has ABEEK goal and semester information. In case of categorical attribute, we suggest a method to quantify them by using binarization. The number of clusters used for K-means clustering, we applied Elbow method using R-squared value representing the variance ratio that can be explained by the number of clusters. The performance of the suggested method was verified by using a log of student registration data from an 'A university' in terms of the simplicity and fitness, which are the typical performance measure of obtained process model in process mining.

Keywords : Curriculum Mining, Registration Patterns, K-means Clustering, Fitness, Simplicity

1. 서 론

대학교를 비롯한 모든 교육 기관은 교육을 통하여 이루고자 하는 것을 정의한 교육 목표를 가지고 있다. 또한 교육기관은 그 교육 목표를 달성하기 위하여 교육 내용을 체계적으로 조직하고 전체적으로 계획한 교육과정에 따라 운영되고 있다. 그러나 대학교에서는 정형화된 교육과정을 따르는 중등 교육기관과는 달리 학생들에게 수강 교과목을 선택하고 교육과정을 이수하는 과정에 자율성을 부여하고 있다. 따라서 교육과정의 흐름을 제대로 파악하지 못하는 학생들은 오히려 설계된 교육과정에서 벗어나는 문제가 발생하고 있다.

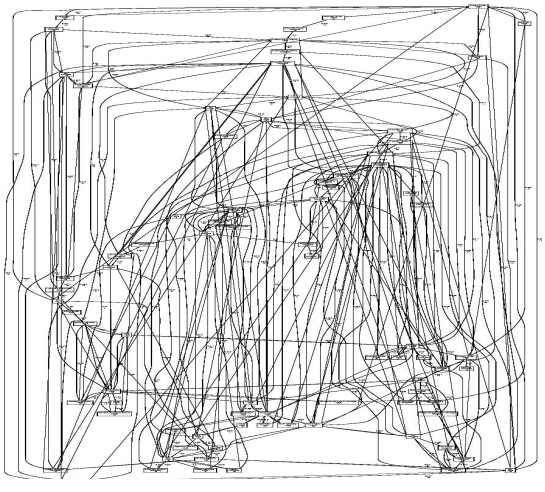
이러한 문제는 공학 계열 교육과정의 경우 2000년대 초반에 공학인증 제도가 도입된 이후 더 심화되었다. 같은 전공 내에서도 공학 인증 교육과정을 이수하는 경우에는 체계화된 교육과정이 제시되지만, 공학 인증 교육

Received 10 September 2015; Finally Revised 2 November 2015; Accepted 19 November 2015
† Corresponding Author : choijy@ajou.ac.kr

과정을 이수하지 않는 경우에는 다른 복수전공이나 부전공을 필수로 이수하기 위해 타 전공의 교육과정까지 고려해야 하므로 학생들이 교육과정의 흐름을 따르는데 있어서 더 큰 혼란을 겪게 되었다.

따라서 학생들이 교육과정 이수 과정에서 겪는 혼란을 최소화하며 설계된 교육과정에 충실히 따르게 하기 위한 방법이 필요하며, 이를 위해서는 먼저 학생들의 교육과정 이수 패턴을 분석하고 실제 이수 패턴과 설계된 교육과정 간에 차이가 발생하는 부분을 파악하는 것이 요구된다. 또한 더 나아가서는 이러한 분석 결과를 바탕으로 학생 별 과거 수강 교과목 데이터를 기반으로 관심 전공 분야 별 수강 지도를 체계적으로 할 수 있는 방법이 제안 되어야 한다. 본 연구에서는 이를 위한 방법으로 데이터 마이닝과 프로세스 마이닝 분야에서 연구되고 있는 결과를 교육 분야에 응용한 커리큘럼 마이닝 기술을 적용하고자 한다.

데이터 마이닝이란 컴퓨터 기반 정보시스템의 대용량 데이터 저장소에서 데이터를 읽어 들이고, 정보를 생산하여 지식을 얻어내는 기법이다. 이를 이용하여 교육 현장에서 발생하는 독특한 형태의 데이터에 대해 탐구하는 방법을 개발하고, 학생에 대한 이해와 그들이 학습하는 교육과정에 대한 이해를 깊게 하기 위한 것이 교육 데이터 마이닝이다. 한편 프로세스 마이닝이란 정보시스템을 사용하는 과정에서 기록된 데이터를 바탕으로 프로세스에 대한 유의미한 정보를 찾는 연구 분야로서, 데이터 중에서도 특히 프로세스와 관련된 데이터에 초점을 맞추고 있다. 커리큘럼 마이닝은 프로세스 마이닝을 교육 분야에 응용한 것으로 교육정보시스템에서 발생한 데이터를 이용하여 교육과정(커리큘럼)에 관한 유용한 정보를 찾는 연구 분야이다[1].



<Figure 1> Unstructured Registration Pattern Model(Using Default Option of Heuristics Miner)

본 연구에서는 이러한 방법을 적용하기 위한 선행 단계로서 프로세스 마이닝 소프트웨어인 ProM[21]에서 제공하는 프로세스 마이닝 기법을 이용해 A대학교 산업공학과 학생들의 실제 수강 이력에 대한 수강패턴 모델을 도출하고자 하였다. 그러나 도출된 모델에서는 예외적인 프로세스 동작이 많고, 프로세스 단계 간의 인과 관계가 약하기 때문에 복잡도가 높은 비구조화 프로세스의 특징을 가졌다. <Figure 1>은 ProM을 이용해서 도출한 비구조화 수강패턴 모델이다. 전체 수강 교과목을 대상으로 학생들이 수강한 교과목 간의 모든 패스들 중에 연관관계 계수가 0.9 이상인 패스만을 고려한 디폴트 옵션을 이용해 도출한 수강패턴 모델임에도 불구하고, 매우 복잡한 구조를 보이고 있다. 일반적으로 이러한 비구조화 프로세스는 프로세스 마이닝 기법을 직접 적용하는 것이 매우 어렵기 때문에 프로세스의 구조를 정확하게 파악할 수 없다는 문제점이 있으며 보다 효율적인 프로세스 모델 도출 방법이 필요하다.

본 논문에서는 이러한 비구조화 프로세스 문제를 해결하기 위한 접근 방법으로 군집화 기반 커리큘럼 마이닝 프레임워크를 제안한다. 제안된 프레임워크는 군집화 단계와 프로세스 마이닝 단계의 2가지 단계로 구성된다. 먼저 군집화 단계에서는 교과목이 갖는 속성 값을 기반으로 전체 교과목에 대한 군집화를 수행하여 다수의 교과목을 소수의 군집으로 간결하게 만든다. 두 번째 단계인 프로세스 마이닝 단계에서는 교과목의 군집을 입력값으로 한 프로세스 마이닝을 수행해 구조화된 프로세스를 갖는 모델을 도출한다.

본 연구에서 제안한 방법은 학생의 특성이나 행동에 따라 군집화를 수행하는 기존 연구 방법[1, 2, 4]과 다르게 교육정보시스템에서 획득한 정보를 이용해 학생이 수강하는 교과목 속성 값 기반의 교과목 군집화 방법을 적용한다. 그 이유는 데이터 획득과 정량적 분석의 측면에서 교과목 군집화가 유리하기 때문이다. 학생 관련 데이터는 개인정보로 보호되는 비공개 데이터이므로 데이터의 획득이 어렵고, 학생의 심리상태나 태도와 같은 정성적 특성을 갖는 데이터이기 때문에 군집화와 같은 정량적 분석에 적합하지 않다. 반면, 교과목 데이터는 교육정보시스템 상에 공개 데이터로 데이터 획득이 쉽고, 교과목의 권장 이수학기와 같이 데이터를 정량적으로 파악할 수 있어 정량적 분석이 가능하다.

본 논문의 구성은 다음과 같다. 제 2장에서는 커리큘럼 마이닝에 대한 기존 연구 동향을 소개한다. 제 3장에서는 수강 이력 데이터에 대한 전처리 방법과 군집화 방법을 설명하고, 제 4장에서는 군집화 기반 수강패턴 모델을 도출하고, 수강패턴 모델에 대한 분석을 수행한다. 마지막으로 제 5장에서는 결론 및 향후 연구 과제를 제시한다.

2. 관련 연구 동향

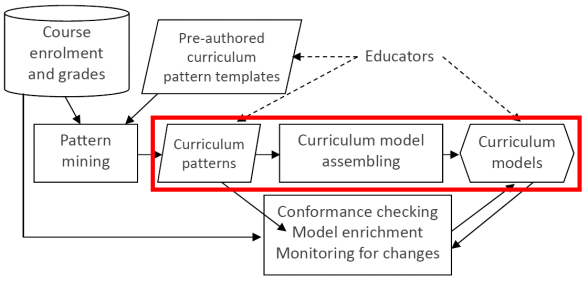
2.1 교육적 프로세스(커리큘럼) 마이닝

프로세스 마이닝이란 정보시스템에 기록된 이벤트 로그를 이용해 프로세스에 관한 유용한 지식을 추출하는 연구 기법이다. 이벤트 로그는 프로세스 수행 과정에서 발생하는 케이스의 집합으로 이루어져있으며, 각 케이스는 이벤트(액티비티)와 발생시각, 작업자 등의 정보로 이루어져있다[3, 17]. 이러한 프로세스 마이닝은 다음과 같이 세 가지 종류로 구분될 수 있다 : 이벤트 로그로부터 프로세스 모델을 생성하는 도출, 도출된 모델과 기존 프로세스 모델을 비교하는 적합도 검사, 기존 프로세스 모델을 개선하여 새로운 프로세스 모델을 생성하는 개선.

교육적 프로세스 마이닝(Educational Process Mining) 혹은 커리큘럼 마이닝(Curriculum Mining)은 프로세스 마이닝을 교육 분야에 적용한 것으로서 교육정보시스템에서 교육에 도움이 되는 지식을 획득하는 것을 목표로 한다. 커리큘럼 마이닝은 2009년에 최초로 Colored Petri Net 형태의 표준 커리큘럼 모델링이 제안된 이후로 프로세스 마이닝의 적용 분야로서 다음과 같이 다양하게 연구 되고 있다[15].

Pechenizkiy et al.[10]에서는 온라인 학습 평가 과정에서의 학생들의 행동을 연구하는데 프로세스 마이닝의 도출, 적합도 검사, 성능 분석 기법을 활용 했고, Southavilay et al.[14]에서는 학생들의 작문 교육 과정에 프로세스 마이닝의 발견 기법을 적용하여 교육 과정 프로세스를 도출했다. Trcka et al.[16]에서는 제약 조건이 있는 교육 이벤트 로그에 대해 프로세스 마이닝의 적합도 검사 분석을 수행했으며, Pechenizkiy et al.[9]에서는 커리큘럼 마이닝이라는 용어를 정립하고 <Figure 2>와 같이 커리큘럼 마이닝의 3대 방식으로 (i) 커리큘럼 모델 도출(학생들의 행동으로 재생산된 완전하고 간편한 커리큘럼 모델 생성), (is) 커리큘럼 모델 적합도 검사(도출된 커리큘럼 모델-이수체계도 간 유사도 확인), (iii) 커리큘럼 모델 개선(정보를 모델에 투영, 특정 교육과정에 대한 이해를 통해 교육과정과 관련된 의사 결정 수행)을 정의했으며, 커리큘럼 마이닝을 구동하기 위한 플러그인을 개발하여 소개했다.

그러나 이러한 기존 연구에서는 교육적 프로세스 마이닝을 수행하는 과정에서 크게 두 가지 문제가 발생했는데, 하나는 도출된 모델이 학생들이 수강한 실제 교육과정 패턴과 잘 맞지 않는다는 문제와 도출된 커리큘럼 모델의 규모가 너무 크고 복잡도가 높아 분석이 매우 어렵다는 문제였다[1]. 일반적으로 커리큘럼 모델과 같이 복잡도가 크고 예외적 행동이 자주 나타나는 모델을 비구조화 프로세스 모델이라고 한다.



<Figure 2> Curriculum Mining Framework

2.2 비구조화 프로세스(Unstructured Process)

비구조화 프로세스 모델은 이해가 어렵고 비구조 모델에 대한 적당한 처리 없이 직접적인 분석이 어렵다는 문제가 있다[18, 22]. 그럼에도 불구하고 비구조화 프로세스는 다양한 연구적 가치가 있는데 그 이유는 현실의 실제 프로세스들은 비구조화 프로세스인 경우가 많고, 비구조화 프로세스의 경우 모델 처리 과정에서 개선의 여지가 많기 때문이다[19]. 프로세스 마이닝 연구 분야에서는 Van der Aalst and Gunth[18]을 통해 비구조화 프로세스 모델에 대해 처음으로 정의했으며 문제 해결 방안으로 이벤트 출현빈도에 따른 필터링과 이벤트 빈도에 따른 군집화를 제안했다.

교육적 프로세스 마이닝에서 비구조화 프로세스 문제 해결을 위해 군집화 기법을 적용한 연구로는 [1, 2]가 있다. Cairns et al.[2]에서는 직업 교육 훈련 분야의 프로세스를 발견하기 위해 훈련생의 특성 치에 따라 군집화를 수행했고, 각 군집 별로 프로세스 모델을 도출해 모델의 복잡도를 낮췄다. Bogarin et al.[1]에서는 대학생들이 온라인 교육 시스템을 이수하는 프로세스를 분석하는 과정에서 군집화를 수행한 모델과 수행하지 않은 모델을 비교했다. 학생 군집화를 하지 않은 모델(기존 모델)과 시험 성적에 따라 직접 학생을 분류한 모델, 학생의 교육 이수 패턴에 따라 EM(Expectation-Maximization) 군집화를 수행한 모델 3가지 모델을 비교하여 모델의 정합도(도출된 모델이 실제를 반영하는 정도)를 비교한 결과 군집화 모델이 기존 모델보다 8.6% 정도 개선된 정합도를 나타냈다.

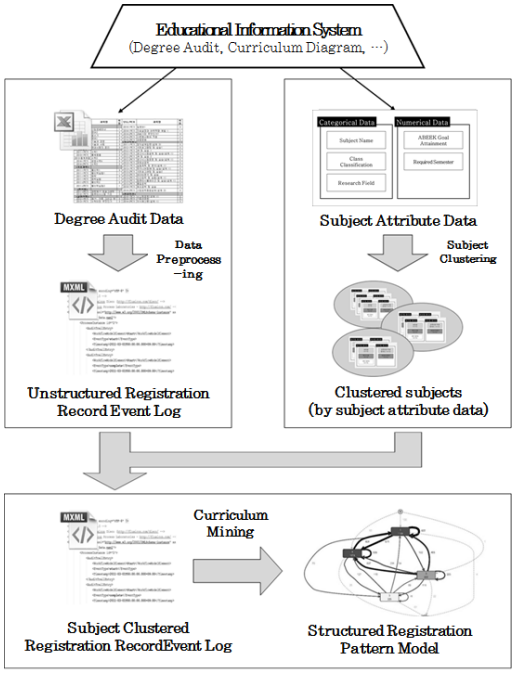
3. 군집화 기반 프로세스 마이닝

3.1 프레임 워크 제안

본 연구에서는 Joo and Choi[6]에 기초하여 <Figure 3>과 같은 군집화 기반 커리큘럼 마이닝 프레임워크를 제안한다. 군집화 기반 커리큘럼 마이닝은 교육정보시스템

에서 수집한 이수 현황표 데이터(학생들의 교과목 수강 이력)과 교과목 속성 데이터(교과목명, 학수구분 등)을 이용해 수행된다.

먼저 이수 현황표 데이터를 이벤트 로그 형태로 변환하는 데이터 전처리를 수행하여 비군집화 이벤트 로그를 생성한다. 본 연구에서는 비군집화 이벤트 로그로 프로세스 모델을 도출하는 경우에 나타나는 비구조화 프로세스 모델 문제를 해결하기 위해 교과목 기반 군집화 방법을 제안한다.



<Figure 3> Suggested Clustering-based Curriculum Mining Framework

교과목 군집화는 교과목 속성 데이터에 K-means 군집화를 적용하여 교과목을 유사한 속성을 갖는 과목끼리 군집화 하는 것을 의미한다. 군집화 된 교과목을 비군집화 수강이력 이벤트 로그에 적용해 '교과목 군집화 기반 수강이력 이벤트 로그'를 도출하고, 이에 대해 커리큘럼 마이닝을 수행하여 비구조화 프로세스 문제를 해결한 '구조화 수강패턴 모델'을 도출한다.

3.2 데이터 전처리

교육정보시스템은 학생들의 수강 이력, 이수체계도나 각 교과목의 강의계획서와 같은 다양한 데이터를 수집하고 제공하여 학생들이 교육과정을 원활하게 이수할 수 있도록 도와주는 역할을 한다[8]. 본 연구에서는 군집화 기반 커리큘럼 마이닝의 첫 단계로써 교육정보시스템의 데이터를 프로세스 마이닝에 알맞은 형태로 바꾸어주는 데이터 전처리 과정을 다음의 두 가지 데이터를 대상으로 수행 한다 : (i) 커리큘럼 마이닝에서 프로세스 도출에 쓰이는 학생의 수강 이력을 나타내는 이수 현황표 데이터와 (ii) 교과목 군집화에 사용되며 교과목의 특징을 나타내는 교과목 속성 데이터.

먼저 이수 현황표는 학생들의 교과목 이수 현황을 기록한 데이터를 의미한다. 본 연구에서는 <Figure 4>와 같은 형태로 저장된 A대학교 산업공학과 123명에 대한 이수 현황표를 이용한다.

Student ID	Year and Semester	Subject Name	Class Classification
1	2011-1	Mathematics 1	Mathematics and Statistics
1	2011-2	Programming for Science	Information System

<Figure 4> An Example of Degree Audit

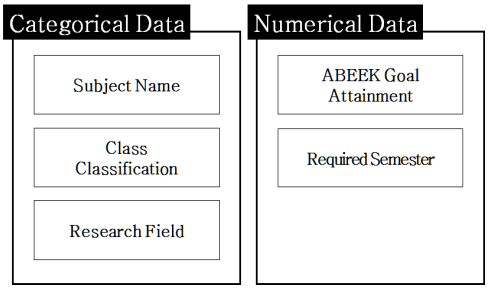
이수 현황표는 익명화된 각 학생의 ID(1에서 123사이의 정수)와 수강 과목, 교과목 이수 시기(년도 및 학기), 각 교과목의 이수 체계 상 학수 구분(전공필수, 전공선택, 수학, 컴퓨터관련 과목) 데이터로 이루어져있다. 데이터 전처리 과정에서는 먼저 이수 현황표를 프로세스 마이닝에 적합한 이벤트 로그 형태로 변환하며, 그 결과를 '수강이력 이벤트 로그(Registration Record Event Log)'라고 정의한다. 수강 이력 이벤트 로그는 <Table 1>과 같이 구성되어 있으며, 각 행은 특정 학생이 수강한 하나의 교과목에 대응된다. 또한 각 열은 학생 식별 번호(ID), 해당 교과목의 이수 시작 시점(학기 시작일자), 이수 종료 시점(학기 종료 일자), 학생이 수강한 교과목을 나타낸다. 수강이력 이벤트 로그를 이용해 바로 수강패턴을 도출할 수도 있지만, 앞에서 언급한 바와 같이 비구조화 프로세스가 도출되는 문제가 발생하므로 수강이력 이벤트 로그의 액티비티인 교과목 속성 기반의 군집화가 필요하다.

<Table 1> An Example of Registration Record Event Log (Preprocessing Completed)

ID	Start Time Stamp	End Time Stamp	Subject
1	2011-03-01 0:00	2011-06-30 0:00	mathematics 1
1	2011-09-01 0:00	2011-12-31 0:00	Programming for Science

교과목 속성 데이터는 교과목을 구분할 수 있는 데이터로서, A대학교 교육정보시스템의 이수체계도와 강의계획서, 한국연구재단의 학술연구 분야 분류표를 통해

수집하였으며, 본 연구에서는 교과목 속성을 <Figure 5>와 같이 과목명, 학수구분, 연구 분야, 공학인증 목표 달성도, 권장이수학기의 5가지 속성으로 구분하였다. 이러한 5가지 교과목 속성은 연속형 데이터(Numerical Data)와 범주형 데이터(Categorical Data)로 나눌 수 있다.



<Figure 5> Subject Attribute Data

연속형 데이터인 공학인증 목표 달성도와 권장이수학기는 연속적인 수로 구성되어 계량화가 가능하다. 특히, 공학인증 목표 달성도는 전문능력, 협업능력, 자주의식의 3가지 세부 속성으로 구분할 수 있으며, 각 속성에 대한 목표 달성도는 0부터 12까지의 정수로 표현된다. 하지만, 과목명, 학수구분, 연구 분야와 같은 범주형 데이터는 그 자체로는 계량화가 불가능하며 단순히 데이터를 구분하는 이름에 불과하다. 따라서 범주형 데이터는 직접 군집화를 적용할 수 없으며 이에 대한 전처리가 반드시 필요하다[5].

범주형 데이터 전처리는 다음과 같이 수행한다. 먼저 본 연구의 대상인 A대학교 산업공학과 123명의 학생들이 수강한 42개 교과목에 대한 과목명은 과목을 식별하는데 사용될 뿐 군집분석의 입력 값으로 사용하지 않는다. 학수구분은 이수체계도 상의 교과목 분류를 의미하며 전공필수, 전공 선택, 수학, 컴퓨터관련 과목 등의 4가지로 이루어져있다. 연구 분야는 각 교과목과 연관된 산업공학의 세부 연구 분야를 의미하며, 본 연구에서는 한국연구재단에서 제작한 학술연구 분야 분류표를 참조하여 총 13개의 세부 분야로 나누었다. 각 연구 분야별 해당 교과목은 <Table 2>와 같다. 본 연구에서는 4개의 학수구분과 13개의 연구 분야에 대해 각각의 교과목의 해당되는 속성을 1로 표시하여 정량화를 실시한다. 예를 들어 <Table 3>은 Creative Design 과목이 기타 산업공학 분야 전공 필수 과목임을 나타내며, Data Analysis 과목은 수학통계 분야의 전공 필수임을 나타낸다. 5가지의 교과목 속성 데이터는 <Table 4>와 같이 정리될 수 있다.

<Table 2> Academic Research Field Classification

Research Field	Subject
E-business	Service Engineering, Product Design
Technology Management	Strategic Management of Technology IP Management
Manufacturing	Production System 1, Production System 2, Facility Engineering
Logistics	Logistics, S.C.M
System Analysis	Engineering Economy, Cost Engineering, System Engineering
Human Factor	Human Factors, Work factor HCI, Usability Engineering
Information System	IE1, IE2, Database, Corporate Solution Information System, Programming for Science, B.C.S
Optimization	OR1, OR2, Optimization
Computer	Manufacturing, Manufacturing Process Control, Automation Systems, Digital Manufacturing
Quality and Reliability	Design of experiment, Quality Engineering, Reliability Engineering
Mathematics and Statistics	Data Analysis, mathematics 1, mathematics 2, Engineering Mathematics, Statistics Application
Probability Model	Simulation
Etcetera Industrial Engineering	Creative Design, Capstone Design, Internship

<Table 3> Categorical Variable Binarization of Subject Attribute Data

Subject	Research Field(13 Attributes)					Class Classification(4 Attributes)			
	E-biz		Math and Stat		Etc. IE	Major Req	Major Sel	Math	Computer
Creative Design	0	…	0	…	**1**	**1**	0	0	0
Data Analysis	0	…	**1**	…	0	**1**	0	0	0

<Table 4> Completed Data Preprocessing Completed

Subject	Required Semester	ABEEK Goal Attainment			Research Field(13 Attributes)					Class Classification(4 Attributes)			
		Expertise	Collabo-ration	Self Conscious	E-biz		Math and Stat		Etc. IE	Major Req	Major Sel	Math	Computer
Creative Design	1	12	9	12	0	…	0	…	1	**1**	0	0	0
Data Analysis	4	6	8	0	0	…	1	…	0	**1**	0	0	0

3.3 교과목 속성 기반 군집화

본 연구에서는 42개 교과목의 군집화를 위한 방법으로 K-means 알고리즘을 적용하며, 군집 개수 K를 결정하는 방법 중 가장 대표적 방법인 Elbow Method를 이용한다[7]. Elbow Method란 <Figure 6>의 식을 사용하여 클러스터 수로 설명되어지는 분산의 비율인 R-Squared 값을 계산하여 클러스터의 개수가 증가함에도 불구하고 설명 가능한 분산의 증가폭이 적어지는 점을 Elbow Point로 판단하고 해당 Elbow Point를 클러스터 개수로 결정하는 방법이다.

W_{SS} : Within Sum of Squares

B_{SS} : Between Sum of Squares

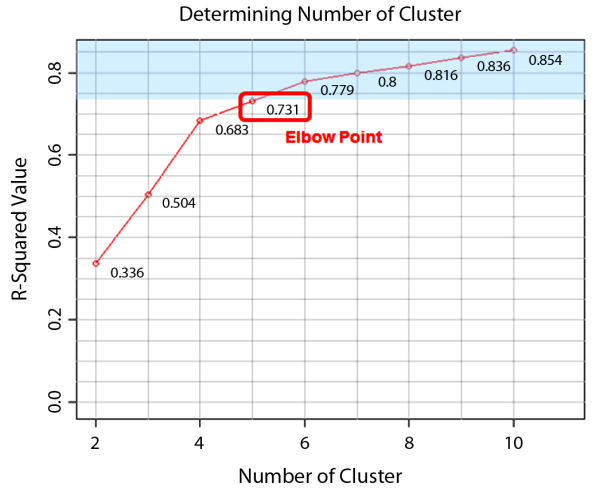
$T_{SS} = W_{SS} + B_{SS}$

$R\ Squared = \frac{Bss}{Tss}$

<Figure 6> Formulation of R-Squared Value

<Table 5> Results of R-Squared Value Calculation

K	Bss	Wss	RSq	Number of Elements for Each cluster
2	551.417	1089.417	0.336	6, 36
3	826.417	814.4167	0.504	16, 18, 8
4	1121.222	519.6111	0.683	9, 10, 18, 5
5	1199.078	441.7556	0.731	9, 10, 10, 8, 5
6	1277.722	363.1111	0.779	10, 9, 5, 3, 3, 12
7	1312.244	328.5893	0.800	6, 7, 5, 3, 10, 3, 8
8	1338.804	302.0298	0.816	7, 4, 3, 8, 3, 8, 6, 3
9	1371.530	269.3032	0.836	3, 4, 9, 3, 5, 4, 5, 7, 2
10	1401.586	239.2476	0.854	4, 7, 2, 3, 5, 5, 3, 4, 6, 3



<Figure 7> Determining Number of Clusters Using Elbow Method

<Table 5>는 군집의 개수에 따라 변화하는 R-Squared 값과 각 군집별 원소 개수를 나타낸다. 이를 그래프로 표현해 보면 <Figure 7>과 같이 군집 수 5를 기준으로 군집 수의 증가에 비해 R-Squared 값의 증가 폭이 급격하게 감소하므로 군집 개수(K)를 5로 결정한다. 이를 이용해 K-means 군집화를 수행한 결과를 공학인증 목표 달성도, 학수구분, 연구 분야에 따라 나타내면 <Table 6>~<Table 8>과 같다.

<Table 6> Cluster Attribute Values for Semester and ABEEK Goal Attainment Level

Cluster	Semester	ABEEK Goal Attainment Level		
		Expertise	Collaboration	Self Conscious
1	0.600	**0.89**	0.26	0.02
2	0.675	**0.55**	**0.52**	0.00
3	0.388	0.17	0.00	0.02
4	0.863	0.00	**0.59**	0.38
5	0.650	**1.00**	**1.00**	**1.00**

<Table 6>은 교과목 군집화 결과 중 이수학기와 공학인증 목표 달성도의 중심(각 속성의 평균값)을 나타낸 것으로, 이수학기의 중심이 1에 가까워질수록 해당 군집에 속한 교과목들은 이수학기가 8학기에 가까워진다는 것(교육 과정 후반기 이수 교과목)을 의미하며, 중심이 0에 가까워질수록 해당 군집에 속한 교과목의 이수학기가 1학기에 가까워진다는 것(교육 과정 초기 이수 교과목)을 의미한다. 공학인증 목표 달성도 또한 그 값이 1에 가까워질수록 해당 군집에 속한 교과목의 학습을 통해 해당 공학인증 목표를 달성할 수 있다는 것을 의미하며, 0에 가까워질수록 목표 달성도가 낮아진다는 것을 의미한다.

<Table 7> Cluster Attribute Values for Semester and Class Classification

Cluster	Semester	Class Classification			
		Major Required	Major Selection	Math	Computer
1	0.600	**0.889**	0.111	0.000	0.000
2	0.675	**0.400**	**0.600**	0.000	0.000
3	0.388	0.100	**0.300**	**0.400**	0.200
4	0.863	0.000	**1.000**	0.000	0.000
5	0.650	**0.600**	**0.400**	0.000	0.000

<Table 7>은 교과목 군집화 결과 중 이수학기와 학수구분의 중심을 표로 나타낸 것이다. 학수구분 속성에서도 중심 값이 1에 가까워질수록 해당 군집의 교과목들은

해당 학수 구분 분류에 속하게 된다는 것을 의미한다.

<Table 8>은 교과목 군집화 결과 중 세부 연구 분야의 중심을 나타낸 것이다. 연구 분야 속성의 중심이 1에 가까워질수록 그 군집에 속하는 교과목들은 대체로 해당 연구 분야와의 연관성이 크다는 것을 의미하며, 중심이 0에 가까워질수록 그 군집은 해당 연구 분야와의 연관성이 낮다는 것을 의미한다.

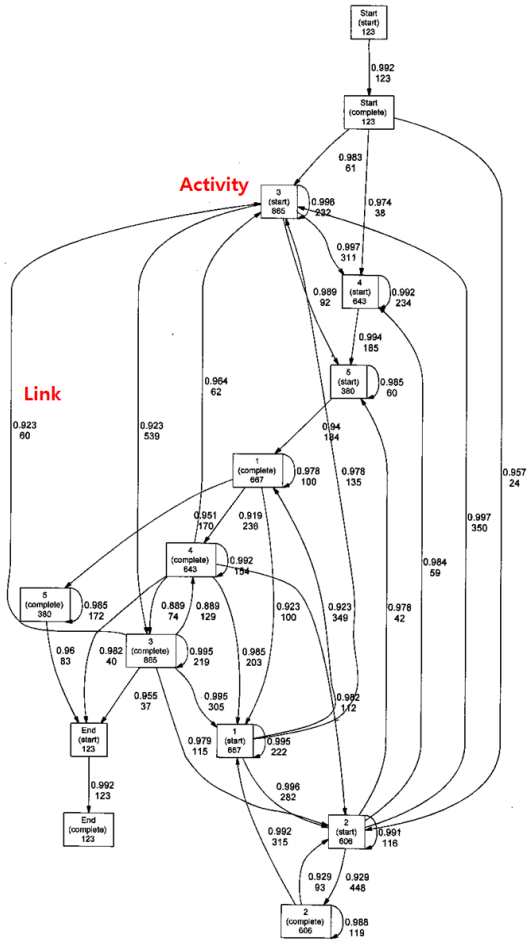
<Table 8> Cluster Attribute Values for Research Field

Research Field	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
E-biz	0.000	0.100	0.100	0.000	0.000
Technology Management	0.000	0.000	0.000	**0.250**	0.000
Manufacturing	0.000	**0.200**	0.000	0.125	0.000
Logistics	0.000	0.100	0.000	0.000	**0.200**
System Analysis	0.000	0.100	**0.200**	0.000	0.000
Human Factor	0.000	0.100	0.000	**0.250**	**0.200**
Information System	**0.333**	0.000	**0.300**	0.125	0.000
Optimization	**0.222**	0.100	0.000	0.000	0.000
Computer	0.111	0.100	0.000	**0.250**	0.000
Quality and Reliability	**0.222**	0.100	0.000	0.000	0.000
Math & Stat	0.000	0.100	**0.400**	0.000	0.000
Probability Model	0.111	0.000	0.000	0.000	0.000
Etcetera Industrial Eng Subjects	0.000	0.000	0.000	0.000	**0.600**

3.4 수강패턴 모델 도출

본 연구에서는 교과목 군집 기반 수강패턴 모델 도출을 위해 ProM 소프트웨어를 이용한다[21]. ProM은 아인트호벤 공과대학교에서 개발한 프로세스 마이닝 소프트웨어로 여러 프로세스 마이닝 알고리즘이 플러그인 형태로 제공되어 프로세스 모델 도출과 도출된 모델에 대한 추가적 분석 기능(적합도 검사, 프로세스 개선 등)을 제공한다. 본 연구에서는 ProM 플러그인으로 제공되는 Heuristics Miner의 기본 Dependency Threshold 설정을 이용해 교과목 군집화 수강이력 이벤트 로그에서 <Figure 8>과 같은 구조화 수강패턴 프로세스 모델을 도출하였다.

사각형(node)으로 표현된 액티비티는 학생들이 각 군집에 속한 교과목을 수강하는 활동을 의미하며, '3(start)'와 같이 '군집번호(Start/Complete)' 형태로 표현된 액티비티는 군집에 속한 교과목의 수강을 시작하거나 종료하는 활동을 표현한다. 이 때, Start 액티비티와 End 액티비티는 각각 입학과 졸업을 의미한다. 액티비티 아래의 숫자는 액티비티의 빈도수를 나타낸다. 액티비티를 연결하는 링크(Link)는 액티비티 간의 선후 관계를 표현한다. 각 링크마다 2개의 속성 값이 부여되어 있는데 위쪽에 소수로 표현된 속성 값은 액티비티 간의 연관 관계를-1에서 1사이의 실수로 나타낸 계수이다. 연관 관계 계수는 1에 가까울수록 링크의 순방향으로 흐르는 연관관계가 깊다는 것을 의미한다. 아래 속성 값은 해당 링크가 이벤트 로그에서 나타난 빈도수를 의미한다. 특히, <Figure 8>은 연관관계 계수가 0.9 이상인 링크만 프로세스 모델에 표현하는 디폴트 옵션을 이용하여 도출한 결과이다.



<Figure 8> Structured Registration Pattern Model(Using Default Option of Heuristics Miner)

4. 수강패턴 분석

4.1 군집별 특성 분석

본 연구에서는 먼저 군집별 속성 값의 중심을 도출한 결과인 <Table 6>~<Table 8>을 이용해 각 군집의 특성을

분석하고, 전체적인 교육과정의 설계 의도를 평가하였다.

군집 1의 이수학기 중심 값을 보면 교육과정의 중기에 수강하는 과목들이 속해있다는 것을 알 수 있다. 공학인증 목표 달성도의 중심 값을 보면 특히 전문능력의 중심 값이 다른 목표 달성도에 비해 매우 높은 값을 가지므로 전문 능력 함양에 중심을 두는 교과목으로 해석되며, 학수 구분의 중심 값을 보면 전공 필수 과목인 것으로 해석된다. 연구 분야의 중심 값을 보면 산업공학에서 주로 다루는 정보시스템, 최적화, 품질 분야 과목들이 주로 속해 있는 것으로 나타났다. 군집 1의 특성을 정의하면 교육과정 중기에 수강하는 산업공학 고유의 전문성을 띠는 전공필수 과목이라고 할 수 있다.

군집 2의 이수학기와 공학인증 교육 목표 달성도의 중심 값은 군집 1과 전반적으로 유사한 양상을 보였지만, 목표 달성도 중 협업 능력과 학수 구분이 군집 1보다 높은 값을 가졌다. 군집 2의 연구 분야는 특정 연구 분야에 국한되지 않으며, 다양한 분야에 걸쳐 있는 것을 볼 수 있다. 따라서 군집 2는 교육과정 후반에 수강하는 과목들 중 전문 능력과 함께 협업 능력을 강조하는 전공과목 군집으로 정의할 수 있다.

군집 3은 5개의 군집들 중에서 이수학기의 중심 값이 가장 작으므로 교육과정상 가장 이른 시기에 수강하는 과목의 군집으로 해석된다. 공학인증 목표 달성도는 전문능력에 대한 달성도가 3가지 핵심 교육 목표 중 가장 높기는 하지만 군집 1이나 군집 2에 비하여 모든 공학인증 목표 달성도가 현저히 낮다. 학수 구분도 전공과목 보다는 수학 과목에 치우치는 경향을 띠며 연구 분야에서도 수학 분야에 속하는 것을 볼 수 있다. 따라서 군집 3은 저학년이 수강하는 수학 분야 과목의 군집으로 정의할 수 있다.

군집 4는 이수학기의 중심 값이 가장 크기 때문에 교육 과정 말기에 수강하는 과목의 군집으로 해석된다. 공학인증 목표 달성도를 보면 전문능력보다는 협업능력과 자주의식을 강조하는 것을 볼 수 있으며 전공 선택과목이 대다수이다. 연구 분야의 중심 값을 보면 기술경영이나 인간공학, 컴퓨터와 같은 분야에 가까운 특성을 갖는다. 즉, 군집 4는 교육과정 후반에 수강하는 전공선택 과목으로써 전문 지식보다는 사회 구성원으로써의 역할을 강조하는 과목의 군집으로 정의된다.

군집 5는 교육과정 중기에 수강하는 과목들로써 특이하게 교육 핵심목표 3가지의 중심이 모두 높은 값을 갖고 있다. 또한 연구 분야의 중심을 보면 기타 산업공학 분야에 치우쳐있는 것을 볼 수 있다. 기타 산업공학 분야는 특정 분야의 지식을 습득하기보다는 종합적인 문제해결 능력을 강조하므로 목표 달성도의 중심 값이 세 가지 모두 크게 나타난 것으로 해석된다. 군집 5는 교육과정 후반기에 학생들이 그동안 배운 산업공학 전문 지식을 이용하여 실생활 문제를 해결하는 것을 목표로 하는 과목으로 이루어진 군집으로 정의된다.

<Table 9> Level of ABEEK Goal Achievement associated with the Semester

Cluster	Semester	Expertise	Collaboration	Self Conscious
3	0.388	0.17	0.00	0.02
1	0.600	**0.89**	0.26	0.02
5	0.650	**1.00**	**1.00**	**1.00**
2	0.675	0.55	0.52	0.00
4	0.863	0.00	0.59	0.38

마지막으로 <Table 9>는 이수학기 순서에 따른 공학인증 목표 달성도를 나타낸 것인데, 이를 통해 A대학교 산업공학과의 전체적인 교육과정 흐름과 설계 의도를 다음과 같이 평가할 수 있다. 교육과정 초, 중기까지는 산업공학 기초와 전문 지식 습득에 관련된 군집 3, 1의 교과목들로 설계되어 있으며, 교육과정 중, 후기로 가면서는 전문 지식 함양보다는 교육과정 초기에 습득한 전문 지식을 바탕으로 사회 구성원으로서의 역할을 준비할 수 있도록 협업능력과 자주의식 배양에 관련된 군집 5, 2, 4의 교과목들이 배치되어 있다. 따라서 전체 교육과정이 시기 별로 학생들이 원활하게 사회에 진출할 수 있는 능력을 함양할 수 있도록 설계되어 있다는 것을 알 수 있다.

4.2 수강패턴 모델 분석

다음 단계로서 [12, 20]에서 제안한 프로세스 모델의 4가지 품질 척도 중 단순성(Simplicity)과 적합도(Fitness)를 기준으로 <Figure 8>의 구조화 수강패턴 모델과 <Figure 1>의 비구조화 수강패턴 모델을 비교하였다. <Table 10>은 비구조화 모델과 구조화 모델의 액티비티와 링크의 개수를 비교한 결과이다. 액티비티 개수는 비구조화 모델이 86개, 구조화 모델이 14개로 약 84% 감소하였다. 링크의 개수는 비구조화 모델이 185개, 구조화 모델이 41개로 약 77% 감소하였다.

<Table 10> Comparison of Structured Model and Unstructured Model Simplicity

Process Model	Number of Activity	Number of Link
Unstructured Registration Pattern Model	86	185
Structured Registration Pattern Model	14	41

모델의 단순성은 일반적으로 프로세스 모델의 액티비티 개수와 링크 개수로 정의되며, 개수가 적을수록 높아지게 된다[20]. 비구조화 수강패턴 모델과 구조화 수강패턴 모델의 단순성을 액티비티의 수와 링크의 수를 이용해 비교한 결과 비구조화 모델에 비해 구조화 모델의 단순성이 크게 증가했다. 모델의 적합도(Fitness)는 도출된 모델을 통해 실제 행동패턴을 얼마나 표현할 수 있을지 나타낸 것이다[11, 12, 20]. ProM을 이용해 계산한 결과 구조화 수강패턴 모델의 적합도는 0.3241(전체 학생의 수강패턴 중 32.41%의 수강패턴을 설명할 수 있음을 의미), 비구조화 수강패턴 모델의 적합도는 0.2892(전체 학생의 수강패턴 중 28.92%의 수강패턴을 설명 가능)으로 나타났다. 결과적으로 구조화 수강패턴 모델은 비구조화 수강패턴 모델에 비해 적합도가 약 12% 정도 개선되었다. 분석 대상 이벤트 로그가 달라 직접적 비교는 어렵지만, 학생 속성 기반 군집화를 수행하여 프로세스 모델을 도출한 기존 연구 Bogarin et al.[1]의 모델의 적합도 개선율 8.6%와 비교해 보면, 본 연구에서 제안한 교과목 속성 기반 군집화가 비구조화 프로세스 문제의 해결에 더 큰 효과를 나타냈다는 것을 알 수 있다.

한편, 비구조화 모델에 비해 구조화 모델의 적합도는 개선되었지만 적합도의 값이 0.3이내로 작은 편이다. 그것은 수강이력 이벤트 로그가 매우 높은 복잡도를 갖기 때문이다. 각 학생은 입학부터 졸업까지 평균 27.7개의 교과목을 수강하며 교육과정을 이수한다. 그 과정에서 매학기 수강 과목이나 수강 순서 결정에 따라 각기 다른 수강패턴을 갖게 된다. 따라서 학생들의 수강패턴이 다양해지기 때문에 일반적인 수강패턴이라고 할 수 있는 공통의 부분이 적어지고, 수강 이력 이벤트 로그를 종합해 도출되는 수강패턴 모델과 실제 학생들의 수강패턴 간의 차이도 커질 수밖에 없었던 것으로 분석된다.

4.3 수강 행동패턴 분석

수강 행동패턴을 분석하기 위해서 <Figure 8>의 수강패턴 모델을 <Table 11>와 같이 각 액티비티 간 링크의 빈도수를 표현한 액티비티 링크 빈도 행렬을 이용해 표현하였다[13]. 행렬의 첫 번째 열은 입학, 군집(1, 2, 3, 4, 5), 졸업 액티비티들의 Start를 나타내고, 첫 번째 행은 입학, 군집(1, 2, 3, 4, 5), 졸업 액티비티들의 End를 나타낸다. 각 셀은 Start 액티비티에서 End 액티비티로의 빈도수를 나타낸다. 예를 들면, 두 번째 행은 입학 액티비티에서 군집 3으로 61회, 군집 4로 38회, 군집 2로 24회의 이동이 발생했음을 의미한다.

<Table 11>을 이용하여 수강 행동패턴을 다음과 같은 방법으로 도출하였다. 먼저 첫 번째 열에서 입학 액티비티를 시작으로 각각의 액티비티에 대해 빈도수가 가장 많은 액티비티를 선택한다. 이 때 동일한 이름의 액티비티는 제외하며, 만일 액티비티가 중복된다면 다음으로 최다 빈도수를 갖는 액티비티를 선택한다. 그 이유는 액티비티의 전체적 흐름을 파악하고자 할 때 입학에서 종료로 진행되는 순방향 링크만 고려하고, 이전에 수행되었던 액티비티로 돌아가는 역방향 링크나 동일 액티비티로 돌아오는 반복 링크를 고려하지 않기 위해서이다.

<Table 11> Activity Link Frequency Matrix

Start \ End	3	1	5	2	4	Graduation
Admission	61			24	38	
3	1050	305	92	115	440	37
1	135	771	170	282	236	
5		184	232			83
2	350	315	42	776	59	
4	136	203	185		388	40

이러한 방법을 적용하여 일반적인 수강 행동패턴을 다음과 같이 도출하였다 : 입학 액티비티에서는 액티비티 3이 선택되고, 그 다음으로는 액티비티 4, 그 다음으로는 액티비티 1이 선택된다. 그러나 액티비티 1에서는 액티비티 4가 최다 빈도이지만 이미 선택되었으므로 그 다음으로 최다 빈도수를 갖는 액티비티 2가 선택된다. 마지막으로 액티비티 5와 종료 액티비티가 선택된다. 선택된 액티비티는 <Table 11>에서 음영 표시되었다. 이를 정리하면, 실제 학생들의 수강패턴이 '입학-3-4-1-2-5-졸업'의 순서를 따르는 것을 알 수 있다. 이러한 결과를 <Table 9>에 표현된 이수체계도와 비교해 보면 액티비티 4와 5의 순서에서 다소 차이가 있었다. 실제 수강패턴에서 액티비티 4는 계획된 이수체계도보다 빨리 수강하려는 경향이 나타났고, 액티비티 5는 계획된 권장 이수시기보다 늦게 수강하려는 경향이 나타났다. 이에 대한 분석은 다음과 같다.

먼저 액티비티 4는 계획된 이수체계도 상에서는 교육과정의 마지막에 수강하는 전공 선택과목들의 군집이다. 하지만 실제 수강패턴에서는 두 번째로 빨리 나타났는데, 그 이유는 요구되는 공학인증 목표 달성도를 통해 알 수 있다. 액티비티 4의 공학인증 목표달성도를 살펴보면 전문능력-협업능력-자주의식이 0-0.59-0.38로 나타났다. 교육과정(이수체계도) 계획 당시 학생들이 전문 지식을 바탕으로 사회인으로써 필수적인 능력을 교육과정 말기에 수강하기를 원했다. 하지만 학생들의 입장에서는 어렵다고 여겨지는 전문능력보다는 협업능력과 자주의식을 요구하는 교과목을 비교적 수월하게 느낀다. 즉 액티비티 4는 학생들의 체감난이도가 적다고 느끼며 교육 이수 과정의 초기에 주로 수강하게 된 것으로 해석 된다.

액티비티 5의 경우 계획상으로는 교육과정 중기에 위치하였지만 실제 수강패턴에서는 가장 마지막에 나타났는데, 그 원인은 액티비티 4의 경우와 반대이다. 액티비티 5는 공학인증 핵심 목표 달성도의 중심이 1-1-1로 매우 높은 수준의 목표 달성도를 전 분야에 걸쳐 요구하고 있다. 즉, 액티비티 5는 모든 부분에서 높은 수준의 달성도를 요구하기 때문에 액티비티 4와는 반대로 학생들의 체감 난이도가 상승하여 수강을 최대한 마지막으로 미루는 것으로 해석 된다.

5. 결 론

본 논문에서는 프로세스 마이닝의 응용 분야 중 하나인 커리큘럼 마이닝에서 학생들의 실제 수강 이력을 바탕으로 수강패턴 모델을 도출하는 문제를 고려하였다. 그러나 수강패턴 모델은 다른 일반적인 프로세스 마이닝 모델과 달리 학생들이 독립적으로 교과목을 신청하는 경우가 많기 때문에 액티비티 간 인과관계가 약해 이해가 어렵고 복잡도가 높은 비구조화 프로세스의 특징을 갖는다는 문제가 있다.

본 논문에서는 이러한 비구조화 프로세스 문제를 해결하기 위해 교과목 속성 기반의 군집화 방법을 기반으로 한 군집화 기반 커리큘럼 마이닝 프레임워크를 제안하였다. 교과목 속성 기반 군집화 방법은 기존의 학생 속성 기반의 군집화와는 달리 속성 데이터의 획득이 용이하며, 정량적인 분석이 가능하다는 장점을 갖는다. 제안된 프레임워크를 K-means를 이용한 군집화를 이용하여 A대학교 산업공학과 학생들의 데이터에 적용해 본 결과 기존의 학생 속성 기반 방법보다 4% 정도 더 높은 적합도를 갖는 프로세스 모델을 얻을 수 있었다.

본 연구에서는 현상을 표현하는 수강패턴 모델을 도출하였는데 향후에는 학생들의 수강패턴을 바탕으로 미래 시점에서 학생들의 수강패턴을 지도하는 방법에 대한 연구가 필요할 것으로 보인다. 또한 본 연구에서는 교과목 속성 데이터의 범주형 데이터를 직접 처리하지 못하는 K-means 군집화 기법을 사용하여 속성 데이터의 전처리를 수행하였지만, 향후 연구에서는 범주형 데이터를 직접 처리할 수 있는 군집화 기법을 도입한다면 데이터 전처리 과정이 좀 더 간결해 질 것으로 기대된다.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(NRF-2014R1A1A 2057194).

References

[1] Bogarin, A., Romero, C., Cerezo, R., and Sanchez-Santillan, M., Clustering for improving educational process mining. *Learning Analytics And Knowledge,* 2014, pp. 11-15.

[2] Cairns, A.H., Gueni, B., Fhima, M., Cairns, A., David, S., and Khelifa, N., Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining. Information Mining and *Management Conference,* 2014

[3] Chung, S.Y. and Kwon, S.T., A Process Mining using Association Rule and Sequence Pattern. *Journal of the Society of Korea Industrial and Systems Engineering*, 2008, Vol. 31, No. 2, pp. 104-111.

[4] Dutt, A., Aghabozrgi, S., Ismail, M.A.B., and Mahroeian, H., Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, 2015, Vol. 5, No. 2, p. 257.

[5] Han, J., Kamber, M., and Pei, J, Data mining : concepts and techniques. Netherlands : Elsevier, 2011, pp. 383-403.

[6] Joo, W.M. and Choi, J.Y., Registration Pattern Analysis and Curriculum Improvement using Clustering and Process Mining. Proceedings of the *KORMS/KIIE/ESK/KSIE/KSS 2015 Spring Conference, Jeju*, pp. 3871-3880.

[7] Kodinariya, T.M. and Makwana, P.R., Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 2013, Vol. 1, No. 6, pp. 90-95.

[8] Lee, T.H., Lee, M.J., and Lee, J.C., A Development of an Adviser Tool for the ABEEK Accredited Program using Curriculum Flowchart. *Journal of the Korea Society of Computer and Information*, 2009, Vol. 14, No. 8, pp. 97-106.

[9] Pechenizkiy, M., Trcka, N., De Bra, P., and Toledo, P., Curri M : curriculum mining. *Educational Data Mining*, 2012.

[10] Pechenizkiy, M., Trcka, N., Vasilyeva, E., Van Aalst, W., and De Bra, P., Process mining online assessment data. *Educational Data Mining*, 2009.

[11] Rozinat, A. and Van der Aalst, W., Conformance testing : Measuring the fit and appropriateness of event logs and

process models. *Business Process Management Workshop*, Springer Berlin Heidelberg, 2006, pp. 163-176.

[12] Rozinat, A., De Medeiros, A.K.A., Gunther, C.W., Weijters, A.J.M.M., and Van der Aalst, W., The need for a process mining evaluation framework in research and practice. *Business Process Management Workshop*, Springer Berlin Heidelberg, 2008, pp. 84-89.

[13] Seol, H.J., Kim, C.H., Lee, C.Y., and Park, Y.T., A new approach to structuring the process based on design structure matrix(DSM). *Journal of the Korean Society for Quality Management*, 2009, Vol. 37, No. 3, pp. 39-53.

[14] Southavilay, V., Yacef, K., and Callvo, R.A., Process mining to support students' collaborative writing. *Educational Data Mining*, 2010.

[15] Trcka, N. and Pechenizkiy, M., From local patterns to global models : Towards domain driven educational process mining. *Intelligent Systems Design and Applications 9th International Conference*, 2009, pp. 1114-1119.

[16] Trcka, N., Pechenizkiy, M., and Van der Aalst, W., Process mining from educational data. United Kingdom : Chapman and Hall, 2010, pp. 123-142.

[17] Van der Aalst et al., Process mining manifesto. *In Business process management workshops*, Springer Berlin Heidelberg, 2012, pp. 169-194.

[18] Van der Aalst, W. and Gunth, C., Finding structure in unstructured processes : The case for process mining. *Application of Concurrency to System Design 7th International Conference on IEEE*, 2007, pp. 3-12.

[19] Van der Aalst, W., Process mining : discovery, conformance and enhancement of business processes. USA : Springer Science and Business Media, 2011, pp. 301-317.

[20] Van der Aalst, W., Process mining : discovery, conformance and enhancement of business processes. USA : Springer Science and Business Media, 2011, pp. 98-107.

[21] Van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M. W., Weijters, A.J.M.M., and Van Der Aalst, W., The ProM framework : A new era in process mining tool support. A*pplications and Theory of Petri Nets,* Springer Berlin Heidelberg, 2005, pp. 444-454.

[22] Veiga, G.M. and Ferreira, D.R., Understanding spaghetti models with sequence clustering for ProM. *Business Process Management Workshop*, 2010, pp. 92-103.

ORCID

Woo-Min Joo | http://orcid.org/0000-0001-9817-9536

Jin Young Choi | http://orcid.org/0000-0001-6397-3107