

# 스피어만 상관계수를 이용한 디지털 융합 강의 전략 시스템<sup>☆</sup>

## Digital Convergence Teaching Strategy System using Spearman Correlation Coefficients

이 병 옥\*  
Byungwook Lee

### 요 약

디지털 융합을 위한 교육은 다양한 학문과 기술들이 컴퓨터를 중심으로 융합하는 것이므로 교육 범위와 방법이 매우 상이하다. 따라서 교육 계획과 강의 전략을 정형화하기 어렵기 때문에 개념적인 정보를 제한적으로 추천하는 문제점이 있다. 본 논문에서는 스피어만 상관 계수를 이용하여 교육 계획과 강의 전략을 제시하기 위한 시스템을 제안한다. 이 시스템은 학계와 산업계의 요구를 기반으로 한 정보로부터 강의 전략 연관성을 찾아 서열화하고, 사용자의 상황과 특성에 적합한 강의 전략 정보를 목록으로 제공하여, 제한적인 개념적 정보 추천의 단점을 해결한다. 성능 실험은 기존의 서비스 시스템들과 비교하여 효과성을 측정하여 정확도와 재현율로 표현하였으며, 성능 실험 결과 정확도는 90.4%, 재현율은 77.6%로 나타났다.

### ABSTRACT

Since educating digital convergence is to unite various sciences and technologies with computer as the central figure, it has different range and methods of education. Therefore, it has problems with recommending limited conceptual information because of difficulties to standardize education plan and teaching strategies. In this paper, I propose education plan and teaching strategy system by using Spearman correlation coefficients. This system is to find a solution against disadvantage of recommending limited conceptual information by ranking relations of teaching strategies from the information based on the demand of industrial and academic fields, and then provides lists of teaching strategy information suitable for user's atmosphere and characteristics. Performance test is to compare effects of precision and recall with existing service systems. The test shows 90.4% of precision and 77.6% of recall.

☞ keyword : Digital convergence, teaching strategy, Spearman correlation coefficients, 디지털 융합, 강의 전략, 스피어만 상관계수

## 1. 서 론

우리나라의 IT(information technology)산업은 혁신적인 기술 발전을 거듭하여 전 세계적으로 시장을 확대하고 주도하였으나 최근에 이르러 새로운 전기를 맞고 있다. 산업이 발전하면서 기술이 고도화되고 경제가 크게 성장하였으며 이에 따라 고객의 요구가 크게 증대하였다. IT관련 제품도 기술 수준만 높게 요구하는 것이 아니라 디자인의 예술적 감성과 수익 모델도 함께 요구하고 있다. iPhone과 iPad를 앞세워 시장 지배를 가속화하

고 있는 애플의 특징은 디지털 융합을 적극적으로 개척한 사례이다. 앞으로의 IT 제품은 기술만으로 성장할 수 없으며 예술적 감성과 창의성과 다양한 인문학적 지식과 수익 모델 등이 조화롭게 반영되어야 소비자의 반응을 얻을 수 있다. 이러한 사회적 요구에 부응하기 위해서 대두된 학문이 디지털 융합이다. 대학에서는 디지털 융합을 적극적으로 추진하고 적용할 수 있는 인재를 양성해야 한다.

2000년대 초에 ACM<sup>1)</sup>에서는 디지털 융합을 포함하는 교과과정에 대한 연구결과들을 발표하고

\* 종신회원 : 경원대학교 IT대학 교수  
leebw@kyungwon.ac.kr

[2010/09/06 투고 - 2010/09/08 심사 - 2010/11/10 심사완료]

1) ACM(Association for Computing Machinery): 1947년 설립된 대표적인 컴퓨터 분야 국제학회.

☆ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

각 대학에 제안하였다[1][5]. 이에 따라 미국의 대학들은 디지털 융합 과목을 적극적으로 교과과정에 반영하고 있다. 사회에서 요구하는 제품은 여러 가지 기술과 예술과 인문학과 사회과학적인 지식을 통합적으로 요구하고 있으나 국내에서는 이에 대한 연구와 준비가 미비하여 지금까지 디지털 융합을 교과과정에 충분히 반영하지 못하고 있다. 공학 인증을 수행하는 대학의 학과에서는 공학 윤리를 교과과정에 반영하고 있는 것이 대표적인 디지털 융합과 관련된 교육 사례이다.

사용자의 개별적인 선호에 따른 정보를 추천하여 서비스를 제공하는 인식 기술은 입력에 할애되는 시간이 상당히 많이 소요되며, 효율적인 정보 제공을 위한 정보 검색 기법은 사용자의 개별적인 취향과 선호에 따른 정보를 추천하기보다 개념적인 정보를 제한적으로 추천하게 되어 검색을 위한 시간이 필요하게 되므로, 사용자의 상황과 특성에 따른 교육 계획과 강의 전략 정보 선호도를 제공하기 힘들고, 사용자가 원하는 정확한 정보를 추천하기가 어려우며, 개념적인 정보만을 제한적으로 추천한다[15].

본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 교육 계획과 강의 전략을 추천하는 시스템을 제안한다. 적은 정보와 연관성을 스피어만 상관 계수를 이용하여 서열화하고 사용자의 상황과 특성에 적합한 정보를 서열화한 목록으로 제공하므로 시스템에서 존재하는 제한적인 개념적 정보 추천의 단점을 해결하였다. 제안된 방법의 수행 과정에서는 특성을 추출하여 특성 처리와 특성 클러스터링 처리를 수행하고 스피어만 상관 계수를 이용하여 예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들의 정보를 기반으로 사용자의 상황별 선호도를 예측하여 유사 선호도를 구하고 유사 선호도로부터 서열화된 최종 추천 강의 전략 객체 목록을 생성하여 제공한다.

성능 실험은 기존의 서비스 시스템들과 비교하여 효과성을 측정하여 정확도와 재현율로 표현하

였으며, 성능 실험 결과 정확도는 90.4%, 재현율은 77.6%로 나타났다.

## 2. 관련 연구

### 2.1 교과과정 연구 사례

ACM은 컴퓨터관련 학과들을 교육 내용에 따라 CS(Computer Science), CE(Computer Engineering), IT(Information Technology), IS(Information System) SE(Software Engineering) 등의 5개 분야로 분류하고 주기적으로 분류 학과별로 교과과정을 제안하고 있다. 이들 교과과정들은 각각 CS는 2008년, CE는 2004년, IT는 2008년, IS는 2010년에 각각 발표되었다[1][5]. ACM의 CS 부문의 교과과정에서는 2001년 보고서부터 컴퓨터 관련 학과목을 14개의 그룹으로 (표 1)과 같이 분류하고 있다[3][4].

(표 1) CS의 14개 핵심 그룹

학과목 이름
Discrete Structure
Programming Fundamentals
Algorithms and Complexity
Architecture and Organization
Operating Systems
Net-Centric Computing
Programming Languages
Graphics and Visual Computing
Intelligent Systems
Information Management
Software Engineering
Computational Science
Human-Computer Interaction*
Social and Professional Issues*

\* 비교적 최근에 추가된 그룹

ACM의 컴퓨터 관련 학과의 학부 교과과정에

서는 디지털 융합의 교육 내용을 (표 2)와 같이 제안하고 있다. (표 2)의 학과목들을 살펴보면 CS는 CE의 내용과는 91% 유사하고 IT와는 45% 정도 유사하며 IS와는 36% 정도 유사하다. 본 연구에서는 CS 관련학과의 디지털 융합을 제안하므로 CS 교과과정을 중심으로 고찰하고자 한다.

(표 2) CS, CE, IT, IS의 융합 교육 내용

CONTENTS	CS	CE	IT	IS
History of Computing	○	○	○	
Social Context	○		○	○
Analytical Tools	○	○		
Professional Ethics	○	○	○	○
Risks	○	○		
Security Operations	○	○		
Intellectual Property	○	○	○	○
Privacy and Civil Liberties	○	○	○	○
Computer Crime	○	○		
Economics of Computing	○	○		
Philosophical Frameworks	○	○		
Public policy		○		
Legal issues in Computing			○	○
Professional Communications			○	○
Teamwork Concepts and Issues			○	○
Organizational Context			○	
Creativity				○
Business Performance				○

디지털 융합은 1990년대 이후에 컴퓨터학과에서 개발된 학과목으로서 컴퓨터를 배우는 학생들이 기본적인 문화, 사회, 법, 윤리 등의 주제들을 이해할 필요가 있어서 시작되었다[7]. 디지털 융합의 학습목표는 컴퓨터학을 공부한 학생들이 사회적으로 전문적인 주제들에 대한 이해를 충분히 갖추므로써 소프트웨어 개발에 적극적으로 활

용할 수 있는 능력을 갖추는데 있다. ACM에서는 학과목의 학습 목표를 구체적으로 달성하기 위하여 기억, 이해, 응용, 분석, 평가, 창조의 6단계의 목표(표 4)를 설정하고 이를 달성할 수 있는 교과과정과 과목 개발 방법을 제안하고 있다.

ACM의 여러 분과에서 제시하고 있는 학과목들과 인지과학을 구성하고 있는 학과목들과 관련 학계와 산업계에서 요구하는 학과목들 중에서 선별하여 이론과 실습을 연계할 수 있도록 학습 내용을 구성한다[9][10]. 실례로 카네기멜론 대학 등의 전산학과에서는 (표 3)과 같이 전공 선택 과목으로 여러 강좌의 디지털 융합 교육을 실시하고 있다.

(표 3) 국내외 대학의 융합 교육 사례

대학	과목 이름
카네기멜론대	Music system and Information Processing
	Computational Photography
	Computational Neuroscience
	Human Language Technology
MIT	Ethics and the Law on the Electronic Frontier
	Inventions and Patents
	Introduction to Copyright Law
	Management in Engineering
	Computational Biology
	Quantitative Physiology
	Molecular, Cellular and Tissue Biomechanics
서울대	컴퓨터융합응용
	정보통신융합
	IT와 금융공학의 융합
	인터넷 윤리

## 2.2 사용자 추천 서비스 시스템

추천 서비스 시스템은 자동화된 정보 필터링 기법을 적용하여 사용자의 맞춤 정보를 추천 서비스하며 선호도 추천 시스템은 사용자에게 개인

신상, 관심 분야, 선호도 등의 정보를 입력받아 사용자의 선호도에 알맞은 정보를 추천 서비스한다 [12][16]. 선호도 추천 서비스 기반의 시스템은 데이터를 이용하는 방법에 따라 분류하며 시스템에 저장되어 있는 기존 데이터와 사용자의 입력 데이터, 기존 데이터와 입력 데이터를 조합하여 형성된다[14]. 사용자 선호도 추천 서비스 시스템에는 단일 협업 기법과 내용 기반 기법, 지식 기반 기법, 유용성 기반 기법 등이 있으며, (표 4)와 같이 추천서비스 기법을 분류할 수 있다.

(표 4) 추천 서비스 기법의 종류

서비스 기법	기존 데이터	입력 데이터	처리
단일 협업	각 항목에 대한 점수	각 항목에 대한 점수	특정 고객이 부여한 점수를 토대로 유사 고객 파악
내용 기반	각 항목의 특징	각 항목에 대한 점수	각 항목에 대한 사용도와 점수에 따른 분류자 도출
지식 기반	고객의 욕구를 충족시킬 수 있는 항목에 대한 지식	고객의 필요와 흥미에 대한 설명	고객의 욕구를 충족시키는 항목 추정
유용성 기반	각 항목의 특징	고객이 선호하는 각 항목의 유용성	유용성에 따라 각 항목을 분류한 후 선택

단일 협업 기법은 사용자의 기초 정보와 고객들의 선호도를 선별하여 비슷한 패턴을 보이는 고객들을 하나의 그룹으로 묶어 추천 서비스를 결정하며 기존 데이터와 입력 데이터를 각 항목에 대한 점수로 표현하고  $k$ -최대 근접 방법을 이용하여 점수가 유사한 고객을 찾아 처리하므로 기존 데이터의 의존도가 높으며 사용자 상황을 고려하기 어려운 단점이 있다[11][15]. 내용 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징으로부터 사용도와 입력 데이터의 점수에 따라 클러스터링을 이용하여 분류하므로 클러스터링 방법에 의존해야 한다. 지식 기반 기법은 고객 요구에 충족하는 지식의 기존 데이터의 필요와 흥미에 대한 입력 데이터를 이용하여 욕구를 충족시키는 항목을 추정하므로 확률에 의존하게 되

고 기존 데이터가 많이 필요한 단점이 있다. 유용성 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징과 입력되는 데이터로부터 선호 항목의 유용성을 클러스터링 방법을 이용하여 분류한 후 선택하므로 클러스터링 방법에 의존해야 한다[15].

### 2.3 라이브러리 시스템

라이브러리는 구성요소의 접근, 탐색, 제어, 보안에 용이한 포괄적인 데이터베이스이며, 사용자가 구성요소를 생성, 편집, 검증, 합성하기 위한 기능을 제공하며, 확장이 용이해야 한다. 유사한 어휘의 선택은 분류 문제이며, 유사성의 정도는 어떻게 집합(collection)을 조직하느냐에 따라 정해진다. 그러므로 집합의 조직과 선택은 이 모델에 있어서 중요하다[16][17]. 어휘의 관련성을 부여하는 방법에 따라 enumerative와 facet 방법[15], 어휘 처리 수준을 포함하는 분류 방법[16], 그리고 구문 및 의미 분석을 포함하는 방법[17] 등이 있다. 어휘 처리(lexical processing)를 이용한 검색 시스템은 어휘 친화력(lexical affinity) 개념과 정보의 양에 기초한 분류 방법을 사용하여 자연어 문서로부터 자동으로 추출된 속성에 따라 부품을 분류하며, 단일 키워드 기반 시스템보다 향상된 정확도를 가진다.

구문 및 의미 처리(syntactic and semantic processing)를 이용하는 시스템은 LaSSIE(Large Software System Information Environment)이며, 지식 베이스, 의미 분석 알고리즘, 그래픽 브라우저, 사용자 인터페이스로 구성되어 있다. 또한, 라이브러리는 구성요소와 연계된 정보를 관리하고, 구성요소의 등록과 삭제, 구성요소에 필요한 사항의 수집과 관리 등에 관한 사항도 포함해야 한다. ‘라이브러리의 예로는 Raytheon Company의 라이브러리, AT&T Pacific Bell사의 C언어 라이브러리 등이 있으며, 이외에도 Prieto Diaz의 라이브러리, EIFFEL, LaSSIE, AIRS, ROSE사 등에서 구축한 라이브러리가 있다.

형상형성 제어는 여러 리비전이나 형상의 공존

을 허용할 수 있어야 한다. 이런 지원을 위해서는 각 문서에 이름뿐만 아니라 리비전이나 형상을 구별할 수 있는 레이블을 지원해야 하며, 형상관리하는 형상  $C$ 가 문서  $D_i$ 에 의존적인 모든 문서들이 새로운 리비전을 갖게 되었을 때  $D_i$ 의 새로운 리비전을 생성할 수 있어야 한다. 이는 문서에 영향을 주는 모든 전위 문서들이 이 리비전을 공유할 수 있는가에 대한 무결성을 점검함으로써 얻을 수 있다. 형상 형성 제어는 유도된 새로운 리비전들이 같은 전위 리비전으로부터 유도되었는가를 식별할 수 있어야 하며, 주어진 특정한 형상을 구성하기 위해서 사용되는 문서들 각각의 정확한 리비전들을 찾을 수 있어야 한다.

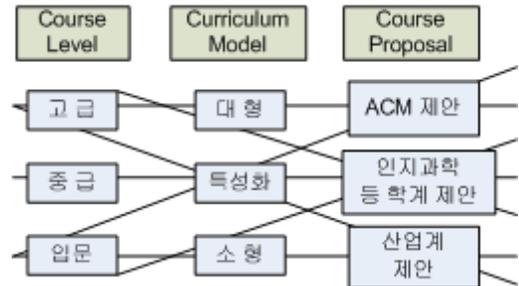
### 3. 시스템 모델

#### 3.1 디지털 융합을 위한 교과과정 설계

디지털 융합을 위한 교과과정은 ACM의 권고에 따라 다음의 원칙을 적용하여 설계한다[1].[5].

- 독립된 과목으로서 전산학의 완전성과 특징을 반영한다.
- 기술 변화에 신속하게 반응하고 학생들을 고무한다.
- 구체적인 설계는 달성하려는 결과에 의하여 유도되어야 한다.
- 상위 과목에서 적용할 수 있는 일관성을 유지해야 한다.
- 도전적인 문제를 해결하기 위하여 최고의 경험을 제공할 수 있어야 한다.

교과과정은 교과과정 모델과 과목의 수준을 설정하고 이에 따라 구체적인 교육 목표를 설계한다. 교과과정 모델은 크게 대형 모델, 특성화 모델, 소형 모델 등으로 구분한다[2]. 대형 모델은 대규모 학과에서 다수의 교수진을 확보하고 있는 경우에 광범위하고 깊이 있게 교육할 수 있는 교과과정이다. 특성화 모델의 특징은 교수진이 많지



(그림 1) 교과과정 설계 기준

않은 경우에 전문분야를 특성화하여 집중적으로 교육하는 모델이다. 소형 모델은 학과에 5-6명의 적은 교수진에 적합하게 교육 범위를 소형화하여 설계한 교과과정이다. 본 연구에서는 우리나라의 학과들이 소규모인 것을 감안하여 특성화 모델과 소형 모델을 기준으로 교과과정을 설계한다[1].

디지털 융합을 특성화하는 특성화 모델 경우에는 입문 과정에서부터 중급과 고급 과정까지 디지털 융합 과목들을 반영하여 (그림 1)과 같이 나타난다. 대형 모델과 소형 모델에서는 디지털 융합을 입문과정을 마친 학생들이 고급 과정에서 활용할 수 있도록 중급 과정에 배정하는 것이 바람직하다. ACM에서 제안하는 중급 과정은 4가지 방식으로 편성된다. 첫째 전통적 주제기반 접근방식(9과목), 둘째 압축 접근방식(5개 과목), 셋째 시스템기반 접근방식(10과목), 넷째 웹기반 접근방식(9과목), 다섯째 혼합 접근방식(앞의 방식을 혼합하는 방식에 따라 8, 7, 6과목) 등이다[8]. 본 연구에서는 우리나라에서 주로 전통적 주제기반 접근방식<sup>2)</sup>을 적용하므로 이 방식을 제안한다.

#### 3.2 학습 목표와 학과목 개발

학습 목표를 설정한 다음에 학과목을 개발한다. 학과목은 과목을 명세하고 설계하고 구현하고 평가하는 절차로 이루어진다. 디지털 융합의 학습

2) A traditional topic-based approach: 알고리즘, 전산기구조, OS, Net-centric 컴퓨팅, 인공지능, 데이터베이스, 디지털 융합, 소프트웨어 개발, 프로젝트.

목표는 ACM의 6단계 학습 목표<sup>3)</sup>를 (표 5)와 같은 기준으로 적용하였다<sup>[2]</sup>. 이론 과목에서 기억, 이해, 응용, 분석 단계를 학습하고 실습 과목에서 응용, 분석, 평가, 창조 단계를 학습하도록 학습 목표를 설정하였다. 디지털 융합은 중급 과정에서 이론 교육을 하고 고급 과정의 관련 학과목들에서 연계하여 실습 교육을 수행한다.

(표 5) 디지털 융합의 학습 목표

단계	범주	인지 처리
1	기억	디지털 융합의 제반 개념
2	이해	제반 개념들을 비교, 분류
3	응용	융합 개념들을 현실에 적용
4	분석	융합 제품들을 차별화
5	평가	융합 제품들을 평가
6	창조	융합 제품 생성

디지털 융합 과목들은 IT 시장과 학계에서 요구하는 사항들을 모두 명세한다. 디지털 융합을 구성하는 교육 내용은 ACM 4개 분야 교과과정 중에서 2 개 이상의 분야에서 추천하는 내용과 가드너<sup>4)</sup>가 제시하는 인지과학의 학과목들과 관련 학계와 산업계에서 요구하는 내용 등으로부터 학과목을 작성한 것이 (표 6)의 학과목들이다. 이들 과목 중에서 학생들의 관심을 충족시킬 수 있는 여부를 검토하여 결정한다.

(표 7)은 여러 제안들이 중복되는 것과 유사한 내용을 분류하고 정리하여 만든 학과목들이다. 강의는 한 학기 동안 주당 3시간씩 이론 강의를 기준으로 작성한다. 인지과학은 범위가 넓어서 신경과학과 인공지능을 별도로 구성하였다. 신경과학에는 뇌과학과 진화론을 포함하며, 인공지능에는 자동기계론을 포함하였다. 윤리와 시민의 자유는

같은 그룹으로 분류하고, 경제와 마케팅도 같은 그룹으로 분류하였다. Review는 시험을 앞두고 충분한 질의와 토론을 통하여 학습을 심화하는 목적으로 설계하였다.

(표 6) 디지털 융합 학과목 추천 목록

제안 부문	학과목
ACM 제안 학과목	History of Computing
	Social Context of Computing
	Professional Ethics
	Privacy and Civil Liberties
	Intellectual Property
	Economics of Computing
	Philosophy
인지과학 학과목	Creativity
	심리학
	신경과학
	인공지능
	철학
관련학회/산업계	언어학
	인류학
	마케팅
	예술

(표 7) 디지털 융합 개론 강의계획서

순서	교육 내용
1	융합 개요
2	컴퓨팅 역사
3	인지과학: 심리학, 철학
4	인지과학: 언어학, 인류학
5	인공지능/자동기계
6	신경과학/뇌과학/진화론
7	review
8	중간고사
9	마케팅/경제
10	지적 재산권
11	컴퓨팅의 사회적인 맥락
12	윤리/시민의 자유
13	예술
14	창의성
15	review
16	기말고사

3) 학습 목표: ACM의 "Computer Science Curriculum 2008" 보고서의 제4장1절1과 On the Nature of Learning Objectives 참조.  
 4) Howard Gardner(1943 - ): 미국 발달심리학자. 인지과학의 구성 학문 제시. 인간의 8가지 복수 지능(언어, 논리-수학, 음악, 공간, 신체-운동, 자연, 대인관계, 내재적 지능) 이론 제시.

### 3.3 상황별 객체 정보 및 클러스터링 처리

다양하게 변화하는 학생 사용자의 정보와 특성을 학생 사용자 데이터로 분류하여 상황에 대한 모든 추천 객체로 이용한다[13]. 학생 사용자의 정보를 분석하고 특성을 인식하여 학생 사용자의 상황에 맞는 추천 가능 범위와 학생 사용자의 특성에 따라 추천 형태를 결정하여 학생 사용자의 위치에서 가까운 범위 내의 사용자의 특성에 적합한 추천 객체를 생성한다.

각 객체의 사후 확률  $P(\omega_1|x)$ 이 가장 큰 값을 가진 객체를 결정하기 위해 사용하였으며 식 (1)과 같이 표현된다[16].

$$P(\omega_1|x) > P(\omega_2|x) \quad (1)$$

특징 벡터들이 포함된 클래스를  $\omega_1$ 로 학과를 나타내고 특징 벡터는  $x$ 로 학생들을 나타낸다. 학생인 특징 벡터가 포함된 학과 클래스의 확률은  $P(\omega_1|x)$ 로 나타내며, 클래스  $\omega_1$ 이 클래스  $\omega_2$ 보다 크면 클래스  $\omega_1$ 이 선택된다. 식 (1)에 대하여 베이저안 정리를 적용하면 식 (2)와 같이 나타난다.

$$\frac{P(x|\omega_1)P(\omega_1)}{P(x)} > \frac{P(x|\omega_2)P(\omega_2)}{P(x)} \quad (2)$$

식 (2)에서  $P(x)$  특징 벡터  $x$ 의 확률 값을 나타내며 결정 규칙에 영향을 미치지 않아 제거하여 특징 벡터를 포함한 학과 클래스 항으로 정리하면 식 (3)과 같이 표현된다.

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \quad (3)$$

특징 벡터를 포함한 학과 클래스  $\omega_1$ 의 확률 값이 학생 특징 벡터를 포함한 학과 클래스  $\omega_2$ 의 확률 값보다 크면 학생 특징 벡터를 포함한 클

래스  $\omega_1$ 의 확률 값이 선택된다.

식 (3)은 우도비(likelihood) 계산을 위한 수식으로 표현되며 확률 밀도 함수의 예측이 가능하고 선호도 예측 처리를 위해 사용하였다.

학생 집단의 특성 클러스터링 처리는 유사한 선호도를 가지는 학생들을 몇 개의 의미 있는 학과 군집으로 나누기 위하여 사용한다. 객체의 각 속성에 대한 선호도를 각기 다른 차원으로 구성하여 좌표 평면의 점으로 표현하고 기존 학생 사용자들을  $k$ 개로 클러스터링 한다. 클러스터링을 통하여 나누어진  $k$ 개의 각 군집의 대푯값들과 테스트 사용자의 각 속성의 선호도 값에 대한 거리를 계산하여 가장 최솟값을 가지는 클러스터를 선택한다. 결정된 클러스터에 속하는 다른 사용자들은 테스트 사용자에게 대한 새로운 클러스터로 구성된다[14][17]. 클러스터링을 위한 알고리즘은 [그림 2]와 같다.

```

Begin
  Initialize the Clustering
  Get Input cluster k
  If cluster k is exist then next
  else goto End
  Clear N
  Add N to Evaluation(n)
  Process the mean and
  Add the Expand(n) to N
  new cluster k calculate
  form collect data to classification
End

```

(그림 2) 클러스터링 알고리즘

클러스터  $k$ 를 결정하여  $k$ 개의 초기 클러스터 중심을 선택하고 일반적으로 주어진 표본 집합의 처음  $k$ 개의 표본을 임의로 선택한다. 각 사용자들을 각 클러스터의 중심과 가장 가까운 거리에 있는 군집 영역에 분배하여 모든 군집에 대하여 해당 군집에 포함된 모든 사용자들의 선호도들로부터 새로운 클러스터 중심을 계산한다. 모든 군집에 대하여 기존의 중심과 새로운 중심의 차이가 없을 때까지 분배를 반복하고 그렇지 않으면

수렴하여 데이터로 분류한다.

### 3.4 스피어만 상관 계수를 이용한 선호도 예측

스피어만 상관 계수는 데이터가 순서에 의한 척도인 경우에 사용된다. 측정형 변수나 순서형 분류형 변수들의 상관관계를 자료의 순위 값에 의하여 계산하는 방법으로 순서형 분류형 변수들의 상관관계를 계산한다. 데이터의 값 대신 순위를 이용하는 상관 계수이며 데이터를 작은 것부터 차례로 순위를 정하여 서열 순서로 바꾸어 순위를 이용해 상관 계수를 구하는 방법이다. 두 클러스터 간의 연관 관계가 있는지 없는지를 밝혀주며 데이터에 이상점이 있거나 표본 크기가 작을 때 유용하게 사용된다[15].

사용자간의 유사도를 계산하기 위하여 사용자 의 객체에 대한 선호도를 테이블로 형성한 후 식 (4)와 같이 스피어만 상관계수를 이용하여 선호도 점수를 변환한다.

$$r = 1 - \frac{\sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \quad (4)$$

변수  $x$ 의  $i$ 번째 관측치의 순위를  $R(x_i)$ 로 나타내고 변수  $y$ 의  $i$ 번째 관측치의 순위를  $R(y_i)$ 로 나타내며 전체 서열 순위는  $r$ 에 의해 구해진 순위로 결정된다.

스피어만 상관 계수를 개선한 알고리즘은 [그림 3]과 같다.

데이터로 구성된 클러스터들에 대해 객체 목록을 생성하기 위하여 클러스터링된 데이터들의 연관 관계를 구하고 이를 서열 순서로 구성하여 스피어만 상관 계수를 이용하였다. 최초의 순위를 구하여 1순위로 부터의 상관 계수를 1과 -1로 나타내어 순위를 조절한다. 두 데이터의 순위가 완전히 일치하면 +1로 표현하고 두 데이터의 순위가 완전히 반대이면 -1로 표현된다.

```

Begin
  Initialize the Spearman
  Get Input Vector XI
  If Vector XI is exist then next
  else goto End
  Clear NL
  Add the NL to Evaluation(n)
  Process the phone transition and
  Add the Expand(n) to NL
  Insert BP to BPframe
  Get the frame of word list BP for transition
  Process word transition
  Process the likelihood(ALL n∈CL)
End
  
```

(그림 3) 스피어만 계수를 개선한 알고리즘

### 3.5 라이브러리 관리

본 논문에서는 확장된 facet 방법을 사용하여 라이브러리를 구축한다. 이 방법은 새로운 정보를 추가하기 쉽고 라이브러리 확장에 쉽게 적용된다. 또한, 현재 정보 단위의 계층적 관련성 표현의 자동화는 시스템의 규모와 소프트웨어의 규모가 커짐에 따라 복잡한 문제점을 가지고 있다. 이 문제를 해결하기 위해 본 논문에서는 사용자가 사용자 인터페이스를 통해 본 논문에서 제시한 시스템을 사용할 때의 작업으로 각 컴포넌트들 간의 계층 관계를 관리하여 제한적으로 해결하였다. 컴포넌트가 점진적으로 증가하여 라이브러리의 규모가 증대되면 이 방법은 문제가 발생할 수 있다. 이 문제 해결을 위해 컴포넌트를 검색하기 위한 find() 함수를 구현하여 컴포넌트를 검색하고, 다른 판단 기준을 만족하는 컴포넌트를 찾는 next() 함수를 호출한다. 이는 라이브러리 관리에 이용되며, 내부에 Retrieve() 함수를 두어 재귀적으로 원하는 음소를 탐색하도록 하였다.

라이브러리는 지원되는 형상형성 제어의 Phoneme Retrieve 클래스에 의해 관리되며, 음소 제어를 위한 컴포넌트를 관리할 수 있으며, 형상형성 제어 지원으로 라이브러리에 음소 컴포넌트를 등록, 검색할 수 있는 메소드와 함수를 지원한다. [그림 4]는 Phoneme RetrieveSize 클래스이다.

```

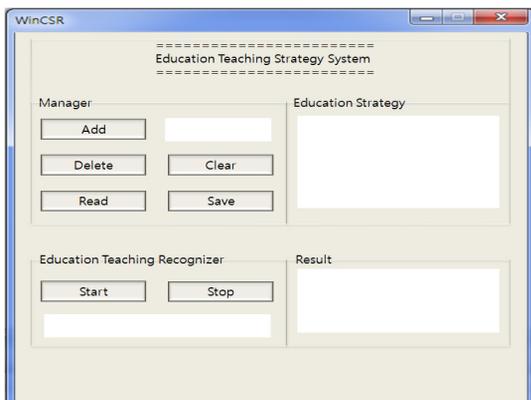
class PhonemeRetrieveSize : public PhonemeRetrieve {
public:
    enum matchType { lessThan, equalTo, moreThan };
    PhonemeRetrieveSize(PtWindowsObject pParent, char
drive, long siz, matchType t);
protected:
    int checkFile(struct fsblk& fs);
    long siz;
    matchType match;
};

```

(그림 4) Phoneme RetrieveSize 클래스 알고리즘

## 4. 평가

본 논문에서 제안한 시스템의 성능을 평가하기 위해 효과성을 측정하여 정확도와 재현율로 나타낸다. 정확도(Precision)는 검색된 데이터들 가운데 연관된 데이터의 비율을 계산하여 연관된 데이터만을 선택하였을 경우 높은 결과를 보인다. 실제 선택한 데이터와 추천한 데이터의 교집합을 추천한 데이터로 나누어 표현한다. 재현율(Recall)은 검색된 데이터를 관련된 모든 데이터의 수로 나눈 수치이다. 선택한 데이터와 추천한 데이터의 교집합을 선택한 데이터로 나누어 표현한다[16]. 실험 데이터는 3장에서 설계한 교과과정 설계, 학습 목표와 학과목 개발을 추천할 수 있도록 하였다.



(그림 5) 탐색 실행

[그림 5]는 시스템에서 처리되는 과정을 보인 것이다.

예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였다. MAE의 대한 수식은 식 (5)와 같다.

$$|E| = \frac{\sum_{i=1}^N |\varepsilon_i|}{N} \quad (5)$$

평균 절대 오차는  $E$ 인 평균 오차에 절대 값에 의해 나타낸다.  $N$ 은 예측하기 위한 총 회수를 나타내고  $\varepsilon_i$ 는 예측 값과 실제 값의 오차를 나타내며  $i$ 는 예측 범위를 표현한다. 식 (5)를 이용하여 예측 값의 정확성 평가를 수행하였으며 (표 8)은 기존 연구와 비교한 결과이다.

(표 8) 예측의 정확성 비교 평가 결과

구분	MAE
단일 협업	0.312
내용 기반	0.301
지식 기반	0.248
유용성 기반	0.227
제안 방법	0.210

제안 방법과 기존 연구의 비교 결과 기존 연구보다 정확성의 예측에서 좋은 예측 결과를 보였다. 성능 평가의 또 다른 방법은 추천 리스트를 평가하는 방법이다. 추천 리스트를 평가하기 위한 방법으로 정확도와 재현율을 사용하고 있으며 정확도는 추천 리스트 중 몇 개의 학과목 개발을 실제로 선택했는지를 평가하고 재현율은 사용자가 선호하는 학습 목표 중 얼마나 많은 학습 목표 추천되었는지를 평가한다. 추천 리스트를 평가하기 위해 정확도와 재현율을 이용하여 기존 연구와 비교 실험을 수행하였다.

디지털 융합을 구성하는 교육 내용은 ACM 4개 분야 교과과정 중에서 2개 이상의 분야에서 추천하는 내용과 가르쳐가 제시하는 인지과학의 학과

목들과 관련학과와 산업계에서 요구하는 내용 등의 학과목들이며 이들 과목 중에서 학생들의 관심을 충족시킬 수 있는지 여부를 검토하여 결정한다. 여러 제안들이 중복되는 것과 유사한 내용을 분류하고 정리하여 만든 학과목들로서 강의는 한 학기 동안 주당 3시간씩 이론 강의를 기준으로 작성한다. 다양하게 변화하는 학생들의 정보와 특성을 학생들이 원하는 데이터로 분류하여 필요상황에 대한 모든 추천 객체로 이용한다. 학생들의 원하는 정보를 분석하고 특성을 인식하여 학생들의 필요상황에 맞는 추천 가능 범위와 학생의 특성에 따라 추천 형태를 결정하여 학생의 필요범위 내의 학생의 특성에 적합한 추천 객체인 학과를 생성한다. (표 9)에서 보이는 결과는 사용자인 학생의 측면과 사업계의 측면을 고려하여 상황에 맞는 추천의 정확도와 재현율로 나타낸 것으로 정확도는 입력값의 의한 결과값의 기대치를 나타내고 재현율 재검색을 실행했을 때 다시 추천 리스트에 나타나는 확률을 표현한 것이다. (표 10)의 비교 실험은 사용자의 선호도 처리 정확도를 k-means 알고리즘과 Pearson 알고리즘을 통해 Spearman의 선호도 처리 정확도와 예측률, 추천 목록 생성률로 비교한 결과이다.

(표 9) 효과성의 비교 실험 결과

구분	정확도(%)	재현율(%)
단일 협업	86.7	80.4
내용 기반	88.1	78.6
지식 기반	84.7	69.4
유용성 기반	89.4	73.4
제안 방법	90.4	77.6

(표 9)의 실험 결과 기존 추천 서비스 시스템들과 비교하여 정확도는 90.4%, 재현율은 77.6%의 성능을 보였다.

k-means를 이용한 방법, Pearson 상관계수를 이용한 방법과의 사용자가 선호하는 대상의 처리율과 선호대상의 정확성, 그리고 추천 목록 생성의

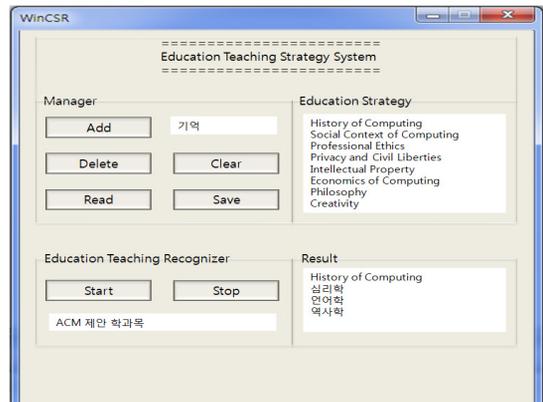
정확성을 각각 선호도 처리, 선호도 예측, 추천 목록 생성으로 표현하였다.

$$F = \frac{2 \times pr \times rr}{pr + rr} \quad (6)$$

비교 실험을 수행하여 (표 9)에 나타내었으며 평가하기 위해 식 (6)과 같이 표현한다. 선호도를 평가하기 위해 F의 수식을 표현하였고 pr은 정확도를 나타내며 rr은 재현율을 나타낸다.

(표 10) 추천 시스템의 비교 실험 결과

구분	k-means	Pearson	Spearman
선호도 처리(%)	94.1	96.8	96.2
선호도 예측(%)	93.3	94.7	95.1
추천 목록 생성(%)	98.5	98.7	98.6



(그림 6) 탐색 실행 결과

실험 결과 선호도 처리에서 k-means를 이용한 방법보다 성능이 높게 나타났으나 Pearson 상관계수를 이용한 방법과는 비슷한 결과를 보였으며 선호도 예측에서는 성능이 가장 높게 나타났다. 추천 목록 생성 비교 실험에서는 세 방법 모두 비슷한 결과를 보였다. [그림 6]은 시스템에서 처리

된 실행 결과를 보인 것이다.

## 5. 결론

2000년대 초부터 ACM에서는 디지털 융합을 포함하는 교과과정에 대한 연구결과들을 발표하고 각 대학에 제안하였으나 이에 대한 연구와 준비가 미비하여 지금까지 디지털 융합에 대한 교육을 교과과정에 충분히 반영하지 못하고 있다. 사용자의 개별 선호에 따른 정보를 추천하여 서비스를 제공하는 인식 기술은 입력 시간이 상당히 많이 소요되며, 효율적인 정보 제공을 위한 정보 검색 기법은 사용자의 개별적인 취향과 선호에 따른 정보를 추천하기보다 개념적인 정보를 제한적으로 추천하게 되어 검색을 위한 추가 시간이 필요하다. 결과적으로 사용자의 상황과 특성에 따른 교육 계획과 강의 전략 정보 선호도를 제공하기 힘들고, 사용자가 원하는 정확한 정보를 추천하기가 어려우며, 개념적인 정보만을 제한적으로 추천한다.

본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 교육 계획과 강의 전략을 추천하는 시스템을 제안한다. 적은 정보와 연관성을 스피어만 상관 계수를 이용하여 서열화하고 사용자의 상황과 특성에 적합한 정보를 서열화한 목록으로 제공하므로 시스템에 존재하는 제한적인 개념적 정보 추천의 단점을 해결하였다.

제안 방법을 수행 과정에서는 특성을 추출하여 특성 처리와 특성 클러스터링 처리를 수행하고 스피어만 상관 계수를 이용하여 예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들의 정보를 기반으로 사용자의 상황별 선호도를 예측하여 유사 선호도를 구하고 유사 선호도로부터 서열화된 최종 추천 강의 전략 객체 목록을 생성한다.

성능 실험은 기존의 서비스 시스템들과 비교하여 효과를 측정하여 정확도와 재현율로 표현하였으며, 성능 실험 결과 정확도는 90.4%, 재현율은 77.6%로 나타났다.

디지털 융합에 대한 교육을 교과과정에 충분히 반영하지 못하고 사용자의 상황과 특성에 따른 교육 계획과 강의 전략 정보 선호도를 제공하기 힘들고, 사용자가 원하는 정확한 정보를 추천하기가 어려우며, 개념적인 정보만을 제한적으로 추천하였다.

본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 교육 계획과 강의 전략을 추천하는 시스템을 제안하여 실험하였다. k-means 알고리즘과 Pearson 알고리즘을 통해 비교 실험을 하였지만 실험 단계에 지나지 않는다고 본다. 따라서 디지털 융합의 대한 교육 교과과정의 설계와 여러 알고리즘을 이용하여 시뮬레이션하고 이에 따른 결과를 교과과정에 직접 반영하여 학생과 사업계에 미치는 영향을 확인하여야 할 것이다.

## 참고문헌

- [1] ACM, IEEE Computer Society, "Computing Curricula 2001 Computer Science - Final Report -" December 15, 2001.
- [2] ACM, IEEE Computer Society, "Computer Science Curriculum 2008: An interim Revision of CS 2001", ACM & IEEE, Dec. 2008.
- [3] ACM, IEEE Computer Society, "Computer Engineering 2004: Curriculum Guideline for Undergraduate Degree Programs in Computer Engineering", ACM & IEEE, Dec. 2004.
- [4] ACM, IEEE Computer Society, "Information Technology 2008: Curriculum Guideline for Undergraduate Degree Programs in Computer Engineering", ACM & IEEE, 2008.
- [5] ACM, AIS, "IS 2010 Curriculum guidelines for Undergraduate degree programs in Information Systems", ACM, 2010.
- [6] 노은하, 박수희, 장준호, "컴퓨터-소프트웨어 분야 수요지향적 교과과정의 개발", 한국컴

- 퓨터교육학회 논문지 제8권 제4호, 한국컴퓨터교육학회, 2005년 7월.
- [7] 이영남 외 5인, “공학 윤리 교육모델 콘텐츠를 이용한 전공과목에서의 공학윤리 교육”, 공학교육연구 제10권 제4호, 2007년 12월.
- [8] 이충기, 홍석원, “웹 기반 프로그래밍 과목의 효과적인 강의전략”, 한국컴퓨터교육학회 논문지 제7권 제3호, 한국컴퓨터교육학회, 2004년 5월.
- [9] 김광수 등저, “융합 인지과학의 프론티어”, 성균관대학교출판부, 2010년.
- [10] 이정모, “인지과학”, 성균관대학교출판부, 2009년.
- [11] 한동조, 박대영, 최기호, “사용자 상황을 이용한 추천 서비스 시스템의 필터링 기법에 관한 연구”, 한국ITS학회 논문지, 제8권, 제1호, 119-126쪽, 2009년 2월.
- [12] 박지선, 김택현, 류영석, 양성봉, “추천시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측 알고리즘”, 정보과학회논문지, 제29권, 제9호, 669-675쪽, 2002년 10월.
- [13] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, “Item-based Collaborative Filtering Recommendation Algorithms”, Accepted for publication at the WWW10 Conference, May, 2001.
- [14] Harry Chen, Sovrin Tolia, Craig Sayers, Tim Finin, and Anupam Joshi, “Creating Context-Aware Software Agents”, Proc. of the first GSFC/JPL Workshop on Radical Agent Concepts, pp.186-200, 2001.
- [15] Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedle, “Application of Dimensionality Reduction in Recommender System-A Case Study”, ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.
- [16] 한학용, “패턴인식 개론”, 한빛미디어, 130-167쪽, 2005년 7월.
- [17] Herlocker, J., “Understanding and Improving Automated Collaborative Filtering Systems”, Ph. D. Thesis, Computer Science Dept., University of Minnesota, 2000.

## ● 저 자 소 개 ●



### 이 병 옥

1973년 연세대학교 공과대학 졸업(학사)  
 1984년 George Washington Univ. 전산학과 졸업(석사)  
 1994년 중앙대학교 전산학과 졸업(박사)  
 1985년- 현재 경원대학교 IT대학 교수  
 관심분야: 분산시스템, 데이터베이스, 디지털 융합  
 E-mail: leebw@kyungwon.ac.kr