

자동분류기반 성격 유형별 도서추천시스템 개발을 위한 실험적 연구*

A Experimental Study on the Development of a Book Recommendation System Using Automatic Classification, Based on the Personality Type

조 현 양(Hyun-Yang Cho)**

〈 목 차 〉

| | |
|---------------------------------|--------------------|
| I. 서론 | III. 도서 자동분류시스템 개발 |
| 1. 연구의 필요성 및 목적 | 1. 시스템 구성 |
| 2. 연구의 내용 및 범위 | 2. 구성 요소별 기능 |
| 3. 이론적 배경 및 선행 연구 | IV. 데이터 실험 및 분석 |
| II. 에니어그램 성격 유형별 특성 및 유형별 선호 도서 | 1. 실험 데이터 및 실험 절차 |
| 1. 에니어그램 성격 유형별 특성 | 2. 실험환경 |
| 2. 성격 유형별 선호 도서 | 3. 성능 측정 결과 및 분석 |
| | V. 결론 및 제언 |

초 록

이 연구의 목적은 개인별 성향이나 성격 유형에 따라 선호하는 도서에 차이가 있음을 전제로, 어린이·청소년을 위한 추천도서의 책소개 정보를 활용하여 개인별 성격유형에 적합한 도서를 합리적으로 추천할 수 있는 서평 자동분류시스템을 개발하는 것이다. 연구에서 사용한 데이터는 국립어린이청소년도서관에서 제공하는 501권의 유아 및 아동도서를 대상으로 하였다. 실험에 활용된 2가지 기계학습 모델(비선형 커널 및 선형 커널) 각각에 대해서 총 6가지의 색인어 가중치 계산 방법과 자질 선택 방법, 그리고 10가지의 자질 선정 임계치 조합으로 구성된 360 개의 분류 모델들을 구성하고 각각의 성능을 측정하였다. 전체적으로는 선형 커널을 이용한 SVM 기반 학습 방법(LIBLINEAR)이 비선형 분류를 지원하는 LibSVM(RBF 커널) 모델보다 더 나은 성능을 보이는 것으로 나타났다. 다만 성능 측정 결과는 뉴스 기사나 논문을 대상으로 한 문헌 분류 성능에 비해서 낮은 것으로 나타났으나, 합리적인 분류 기준이 존재하는 뉴스기사나 주제 분류에 비해서 성격 유형 기반 분류는 그 난이도가 높다는 것을 감안할 때, 초기 실험 결과로서의 의미는 있다.

키워드: 에니어그램, 성격유형, 독서지도, 도서추천시스템, 도서자동분류

ABSTRACT

The purpose of this study is to develop an automatic classification system for recommending appropriate books of 9 enneagram personality types, using book information data reviewed by librarians. Data used for this study are book review of 501 recommended titles for children and young adults from National Library for Children and Young Adults. This study is implemented on the assumption that most people prefer different types of books, depending on their preference or personality type. Performance test for two different types of machine learning models, nonlinear kernel and linear kernel, composed of 360 clustering models with 6 different types of index term weighting and feature selections, and 10 feature selection critical mass were experimented. It is appeared that LIBLINEAR has better performance than that of LibSVM(RBF kernel). Although the performance of the developed system in this study is relatively below expectations, and the high level of difficulty in personality type base classification take into consideration, it is meaningful as a result of early stage of the experiment.

Keywords: Enneagram, Personality type, Reading guidance, Book recommendation system

* 본 연구는 2015학년도 경기대학교 연구년 지원에 의하여 수행되었음.

** 경기대학교 인문사회대학 문헌정보학과 교수(hycho@kyonggi.ac.kr)

•논문접수: 2017년 5월 19일 •최초심사: 2017년 5월 25일 •게재확정: 2017년 6월 3일

•한국도서관정보학회지 48(2), 215-236, 2017. [http://dx.doi.org/10.16981/kliss.48.201706.215]

I. 서론

1. 연구의 필요성 및 목적

독서 환경에의 노출 정도는 개인의 독서량이나 독서율과 연관되어 있다. 독서 환경이란 독서활동이나 능력에 영향을 줄 수 있는 외적인 요인으로, 독서를 장려하는 사회적 분위기나 다양한 독서프로그램과 같은 환경적 요인들을 포함한다. 우리나라 성인들의 독서율이 2011년 66.8%에서 2013년 71.4%로 증가한 요인 중에 하나로 정부의 ‘2012 독서의 해’ 운영 및 독서캠페인 전개, 그리고 공공도서관 이용의 증가 등 독서 진흥을 위하여 진행된 다양한 독서관련 프로그램들의 영향으로 해석해 볼 수 있다.

국가 차원의 ‘2012 독서의 해’ 운영, 지자체별 ‘책읽는 도시’, ‘책의 도시’, 그 외에도 여러 가지 독서캠페인의 전개 등 국민의 독서율과 독서량 증대를 위한 다양한 시도에 힘입어 우리나라 성인 독서율은 2011년의 66.8%에서 2013년 71.4%로 증가하였으나, 독서관련 행사가 지속되지 않았던 2015년에는 65.3%로 다시 감소한 것으로 나타났다. 반면, 학생들의 독서율은 2013년 96%에서 2015년 94.9%로 1.1% 소폭 감소한 것으로 나타났다.¹⁾ 이는 학교도서관에서의 독서동아리를 통한 자발적 독서 참여나 독서 프로그램의 실시와 같은 독서진흥을 위한 다양한 활동으로 성인과 비교할 때 독서 환경에 더 많이 노출되어 비교적 덜 영향을 받았기 때문인 것으로 추측된다.

‘2015 개정교육과정’에서는 독서교육이 강조되고 있으며, 학생들은 독서환경에 더 많이 노출될 수 있는 기회를 가지게 되었다. 개정교육과정의 모든 초·중·고 학생들은 매 학기 국어 수업 시간을 통하여 책읽기, 듣기와 말하기, 쓰기에 이르기까지의 통합 독서교육을 받게 된다. 개정교육과정에서 강조되는 독서의 중요성 인식과 학교에서의 독서교육에 대한 필요성 인지에 힘입어 자녀들의 책 읽기에 대한 부모들의 관심 또한 증가하고 있다. 이러한 부모들의 자녀의 독서에 대한 관심 증가는 가끔 지나친 강요로 작용하여 자녀들의 책에 대한 호감도를 저하시키는 저해 요인이 되기도 한다. 따라서 책에 대한 호감도는 독서에 대한 흥미유발과 직결되는 것임을 고려하여 자녀에게 독서를 강요하기 보다는 자발적으로 참여하도록 유도하는 것이 필요하다.

독서의 효과를 높이기 위하여 독서전문가를 중심으로 읽기 부진아를 위한 독서지도, 강점 강화 독서지도, 독서토론을 통한 독서력 향상 등 지금까지 다양한 교육적 방법이 시도되고 있다. 또한 독서력 향상을 위하여 성격 유형별 상황별, 성인이나 청소년 등 대상별에 따른 맞

1) 문화체육관광부. 2015. 『2015년 국민 독서실태 조사』. 서울: 한국출판연구소.

춤형 독서지도와 같은 다양한 방법들과 목적에 따른 독서지도 방안이 제시되어 왔다.

그러나 독서 효율성 증대를 위해 제시된 지금까지의 방안들은 선정된 책을 읽은 후의 효과를 높이기 위한 방법에 초점을 맞추고 있으며, 적합한 도서의 선정에 대한 중요성은 간과하고 있는 것이 사실이다. 아무리 좋은 책이라 할지라도 개인의 성향에 맞지 않다면 책읽기에 대한 집중력은 저하될 수밖에 없으며, 그 결과 독서의 효과 또한 낮게 나타날 것은 자명한 일이다. 특히 어린이나 청소년들이 일반적으로 선호하는 책도 있지만, 개인별로 선호하는 책은 그들이 처해 있는 상황이나 타고난 성격유형별로 차이를 보인다. 따라서 개인의 성향이나 취향에 맞는 도서의 선정은 독서 효과를 높이기 위한 다양한 독후활동 만큼이나 중요하다.

개인의 성향을 고려한 적합 도서의 선정은 대상의 인성 및 인격 향상에 긍정적인 영향을 준다는 연구결과가 보고되고 있기는 하지만 실제 개인의 성향에 적합한 도서를 선정하는 것은 결코 쉬운 일이 아니다. 대부분의 경우 적합한 도서를 선정하기 위하여 도서관이나 대중매체 혹은 서점에서 제공하는 추천도서를 고려하게 된다. 이러한 추천도서들은 청소년, 아동 등 일반적인 수준에서 적합한 도서를 추천하는 경우로 개인적인 성향에 초점을 둔 맞춤형 추천도서로 보기는 어렵다.

에니어그램은 인간의 선천적 기질과 성향을 구분할 수 있도록 개인별 생각과 행동 유형을 9가지로 분류한 성격유형이론으로 개인별 특성을 고려한 맞춤형 도서의 선택과 선택된 도서를 통한 독서 효율성 증대 방안의 하나로 제안될 수 있을 것이다. 특히 성인과는 달리 어린이나 청소년은 본능적 성향이나 성격 유형의 파악이 용이하여 성향별로 적합한 도서가 존재한다면 맞춤형 도서를 이용한 독서의 효과를 상승시키는 것이 가능할 것이다. 그럼에도 불구하고 개인별 성격 유형이나 심리적 상황에 적합한 도서의 분류를 위한 기준이나 체계적인 도서 자동 추천 방법론은 아직까지 부재한 상황이며, 기존 연구의 경우 연구자의 자의적인 분류기준으로 도서를 선정, 적용하고 있는 실정이다.

특히 학생에게 있어서 독서의 역할과 중요성은 교육현장에서도 강조되고 있으나 입시 중심 교육에 길들여진 학생들은 학과 공부에 대부분의 시간을 투자하도록 요구받고 있으며, 이러한 상황에서 책을 읽을 수 있는 충분한 시간을 소요하지 못하는 것은 당연한 결과이다. 따라서 학생들의 시간을 절약할 수 있으면서도 지적, 정신적 성장을 위하여 필요한 적합한 도서의 선정을 위한 합리적인 방법이 절실히 요구된다.

이 연구의 목적은 상황에 따라 혹은 개인별 성향이나 성격 유형에 따라 선호하는 도서에 차이가 있음을 고려하여 개인에게 가장 적합하고 필요한 도서를 합리적으로 선택할 수 있도록 출판사나 도서관에서 제공하는 서평을 분석하여 개인별 성격유형에 적합한 도서를 추천할 수 있는 자동분류기반 도서추천시스템을 개발하는 것이다.

2. 연구의 내용 및 범위

이 연구는 독서 효과를 극대화 하기 위한 방안의 하나로 개인별 성향이나 성격유형에 맞는 도서를 추천할 수 있도록 자동분류기반 도서추천시스템을 개발하기 위한 것으로, 수행한 연구의 내용 및 범위는 다음과 같다.

첫째, 에니어그램에서 9가지로 구분된 개인별 성격 유형 및 성향에 따라 각 유형별 특성을 대표하는 키워드를 선행연구에 기반하여 추출하였다.

둘째, 국립어린이청소년도서관에서 유아, 초등 저학년, 초등 고학년, 그리고 청소년의 4등급으로 구분하여 제공하는 '사서추천도서' 695권 가운데 유아와 아동용 501권에 대한 책소개 정보를 활용하여 성격유형별 적합도서를 에니어그램 성격 유형을 학습한 3인의 독서전문가가 수작업으로 1차 분류하였으며, 분류된 도서를 최종적으로 에니어그램과 성격유형별 도서 분야의 전문가로부터 검증하는 단계를 거쳤다.

셋째, 다양한 가중치 지정 및 자질 추출 기법을 지원하는 서평 자동분류시스템을 개발하였다.

넷째, 국내에서 출판된 도서 가운데 현재 서점에서 판매중인 약 20만권을 대상으로 저자나 출판사에서 제공하는 서평을 이용하여 도서의 내용을 대표할 수 있는 키워드를 추출하고, 출현빈도와 출현한 키워드 간의 유사도를 측정하여 순위화 하였으며, 이를 기반으로 학습 집합을 구축하였다.

다섯째, 전문가 집단에 의하여 수동 분류된 성격유형별 적합도서의 책소개 정보를 기반으로 이 연구에서 개발한 서평 자동분류시스템을 이용하여 자동분류를 실험적으로 수행하였다.

여섯째, 도서로부터 추출하여 군집화한 키워드와 에니어그램 성격유형별로 대표하는 키워드를 매치시켜 성격유형별로 적합한 도서를 추천하였다.

3. 이론적 배경 및 선행 연구

성격유형별로 적합한 도서를 추천하기 위한 도서 자동분류 시스템과 직접적으로 연관된 연구는 지금까지 진행된 바가 없으나, 문헌정보학이나 전산학 분야에서 지금까지 다양하게 연구되었던 문헌 자동분류 기술을 성격유형별 차이를 고려한 도서 선정과 독서지도에 적용하기 위한 연구가 시도되고 있다.

최형림 등(2003)은 다년간 전자게시판 데이터에 게재된 고객상담 내용을 수집하여 기계 학습 분류기법을 활용하여 담당자를 자동으로 분류, 선정하여 고객의 요구에 신속하게 반응할 수 있는 방안을 제안하였다.

장지숙과 이해영(2009)은 기록의 분류는 기록 자체의 내용보다는 기록이 생산되고 활용

되는 맥락에 초점을 두고 있음을 고려하여 맥락을 중심으로 기록물을 자동으로 분류하기 위한 자동분류시스템을 설계, 제안하였다.

김관준과 이재운(2012)은 국내 주요 학술데이터베이스 검색서비스에서 독서분야 학술지 논문에 부여된 저자 키워드를 기계학습한 결과에 따라 해당 논문들을 재분류함으로써 키워드를 추가로 할당하는 실험을 수행하였다. 그 결과 저자키워드의 재분류를 통하여 디스크립터의 자동 할당이 가능함을 발견하였다.

장재영 등(2012)은 문헌에 나타나는 문법적 요소를 최대한 배제하고 감정단어의 긍정 혹은 부정적 의견을 자동으로 판단하는 자동분류기법인 오피니언 마이닝을 활용하여 영화평론 자동분류기법을 제안하였으며, 실험을 통하여 제안된 기법의 정확도를 평가하였다.

김수아 등(2015a)은 블로그 포스트의 자동 분류를 위해 범주별 고유정보를 반영한 어휘가중치를 제안하였으며, 이를 위하여 범주별로 블로그 문서를 수집하고, 수집한 문서에서 어휘빈도와 문서빈도, 범주별 어휘빈도 등을 고려한 가중치 방법을 제안하였다.

김수아 등(2015b)은 블로그 사이트에서 포스트의 자동분류를 위해 분류별 문서를 수집하고, 수집된 분류별 문서의 어휘빈도와 문서빈도, 그리고 분류별 빈도 등의 다양한 어휘가중치 조합을 통하여 블로그 포스트의 특성에 적합한 가중치 방식으로 TF-CTF-IECDF를 제안하였다.

Russell(2002)는 대학생들을 대상으로 온라인 환경의 교사-사서 합동 수업에서 학생들이 정보를 어떻게 획득, 선택, 활용하는 과정을 관찰하고, 온라인 학습과정에 대한 학생들의 느낌을 인터뷰한 결과 성격 유형별로 정보 추구행태에 차이가 있음을 발견하고, 향후 온라인 환경의 수업을 준비함에 있어서 성격 유형별 특성을 고려하는 것이 필요하다고 주장하였다.

모슬기와 김영숙(2009)은 212명의 중고생을 성격유형별로 구분하고, 학생들의 성격유형과 영어 독해전략 구상에 있어서 어떠한 연관성이 있는지를 살펴 본 결과 성격유형에 따라 독해전략의 활용에 차이가 있음을 발견하였다. 따라서 성격유형별로 선호하는 독해전략을 지도하는 것이 바람직함을 강조하였다.

한윤옥 등(2012)은 초등학교 6학년생과 중학생을 대상으로 에니어그램 성격유형별로 차이점을 관찰하고, 또한 이들과의 심층면접을 통해 성격유형별 독서지도안을 개발하였다. 개발된 독서지도안에 기반하여 성격유형별로 독서지도를 실시한 결과 청소년의 독서태도, 독서성향, 그리고 독서효과에서 유의미한 차이가 있음을 발견하였다. 한윤옥과 조미아(2012)는 어린이와 청소년을 대상으로 성격에 따른 선호도서 목록을 개발하고, 이를 이용하여 독서효과를 극대화하기 위한 독서지도안을 제시하였다.

김국선(2013)은 개인의 성격유형별로 공간구성이나 요구에 대한 선호도에 차이가 있음을 가정하고, 어린이들의 학습능력 향상을 위한 성격유형별 어린이에게 적합한 가구배치 및 공간 구성 방안을 제안하였다.

백진환과 한윤옥(2014)은 학생들의 성격이나 성향을 고려하여 도서의 선정과 독서지도가 이루어져야 독서를 통한 자기 격려와 자기 효능감, 그리고 사회성의 성장에 긍정적으로 작용할 수 있다고 주장하였으며, 이를 초등학교 6학년생을 대상으로 직접 실험하여 검증하였다.

김기열(2015)은 중학생의 성격유형과 기술적 사고 성향간의 관계를 규명한 결과, 선호하는 기술적 사고 성향은 성격유형별로 차이를 보이고 있으며, 성격유형별로 적합한 학습방법을 적용하는 것이 기술적 사고 성향을 향상시킬 수 있음을 증명하였다.

이병열(2015)은 여행업 종사자들의 성격유형과 조직몰입, 고객지향성 및 업무성과에 차이가 있음을 발견하였으며, 직원들의 업무성과 향상을 위하여 성격유형에 따라 차별화된 교육의 필요성을 강조하였다.

조현양 등(2015)은 에니어그램 2번 유형의 초등학교 6학년생을 대상으로 성장지향점 도서를 활용하여 14회에 걸친 독서지도 프로그램을 진행한 결과 정서지능과 자아존중감이 향상되어 독서지도의 효과를 검증하였다.

정덕현(2015)은 글쓰기 수업을 수강하는 대학생을 대상으로 계획단계에서부터 쓰기단계에까지 학생들의 성격유형별로 차이가 있음을 확인하였다. 이에 따라 학생들의 성격유형별 특성을 고려하여 논술문 쓰기교육을 실시한 결과 의미 있는 효과를 발견하였으며, 성격유형별 맞춤교육을 작문지도를 위한 하나의 방법으로 제안하였다.

김경(2016)은 지적장애학생에게 있어서 학업이나 사회적 상황에서 성공을 경험할 수 있는 다양한 기회를 제공하는 것 무엇보다 필요하다고 강조하고, 이를 위하여 이들의 성격유형에 따라 적합한 미술치료프로그램을 포커스 그룹 인터뷰와 다양한 전문가들의 의견을 수렴하여 개발하였다.

지금까지 수행된 선행연구를 조사한 결과 개인별 성격 유형을 분석하여 성격 유형과 상황에 적합한 도서를 추천하기 위한 연구는 아직은 초보적 단계에 있다. 다만, 미국이나 유럽의 경우에도 초중고생을 위하여 학년별로 일괄적으로 추천도서를 선정하여 독서활동을 위한 가이드라인으로 활용하는 정도에 머무르고 있다. 이러한 학교등급별 도서의 추천은 개인의 성격이나 능력을 고려한 맞춤형 도서추천이라기 보다는 표준화된 혹은 보편적인 수준의 도서를 추천하는 것으로 볼 수 있다.

국내의 경우 일부 연구자를 중심으로 상황별 혹은 성격 유형별 도서를 선정하고 이를 소규모의 집단을 총하여 실험적으로 연구를 수행하고 있는 단계이다. 그러나 상황별, 성격 유형별로 적합한 도서를 전문가가 일일이 수작업으로 진행하여 현재 출판되고 있는 도서의 양적 증가를 커버하는 것은 현실적으로 불가능하다. 따라서 다량의 도서가 출판됨과 동시에 독자들에게 추천되어 읽을 수 있는 환경은 결과적으로 대용량의 데이터를 분석할 수 있는 기법을 활용하는 것이 바람직하며, 이를 위한 시도는 아직 초보 단계에 머무르고 있다.

II. 에니어그램 성격 유형별 특성 및 유형별 선호 도서

1. 에니어그램 성격 유형별 특성

에니어그램은 3가지 중심에너지 안에서 다시 3가지의 각 유형으로 분류되어 총 9개의 성격유형에 대해 설명하고 있다. 에니어그램의 9가지 성격유형별 특성을 살펴보면, 협조자 혹은 조력자로 불리는 2번 유형의 성격 소유자는 이타적이지만 소유욕이 강하고 남을 조종하고자 하며, 낙천가로 지칭되는 7번 유형은 활동적이고 열정적인 성격의 소유자로 구분된다. 각 성격 유형별로 별칭과 특성은 다음의 <표 1>과 같다.

<표 1> 에니어그램의 9가지 성격유형별 일반적 특성

| 유형 | 별칭 | 특징 | 심리적 기능 |
|----|-----|------------------------------------|---------------|
| 1 | 개혁가 | 합리적, 이성적, 규범적, 완벽주의적, 편협한 성격 | 합리성, 사회적 책임감 |
| 2 | 조력자 | 협조적, 이타적, 모성애적 유형, 소유욕이 강한 조종하는 성격 | 감정이입, 이타주의 |
| 3 | 성취자 | 성취지향, 실용주의, 자기도취, 자기확신, 야심, 적대적 성격 | 자기존중, 자기개발 |
| 4 | 예술가 | 감수성이 탁월하고 예민한 유형, 개인주의적, 독창적 성격 | 자의식, 예술적 창의성 |
| 5 | 사색자 | 지적, 분석적, 통찰적, 독창적, 공포심이 많은 유형, 괴짜 | 열린 마음, 독창적 사고 |
| 6 | 충성자 | 의무적, 전통준수, 책임감, 의존적 성향, 준비하는 성격 | 동일시, 사회협력 |
| 7 | 낙천가 | 활동적, 개방적, 열정적, 낙천적 성격 | 정열, 실용적 행동 |
| 8 | 지도자 | 강력하고 자기주도적, 자기신념, 단호하고 독재적인 성격 | 자기주장, 지도성 |
| 9 | 중재자 | 태평, 냉정하고, 수용적이고 믿음직함, 수동적 성격 | 수락성, 수용성 |

(출처: 윤운성 2001)

에니어그램의 9가지 성격 유형별로 그 특성이 긍정적 측면으로 발전할 경우와 기질이 잘못 발휘되어 부정적으로 나타날 경우로 나누어 볼 수 있다. 예를 들어, 3번 유형의 성격을 가진 사람은 긍정적 측면에서 목표 지향적이고 도전적이어서 매사에 적극적인 자세를 보이지만 부정적인 성향으로 발전한다면 목표달성을 위하여 지나친 경쟁의식을 갖게 되고, 그로 인하여 다른 사람들과 적대적 관계를 보이는 경우가 발생한다. 각 성격 유형별 긍정적 특성과 부정적 특성은 다음의 <표 2>와 같다.

에니어그램의 성격 유형의 특성에 따른 개인별 성향에 차이가 있다면, 성향이 다른 개인은 취미가 다른 것처럼 선호하는 책의 종류도 다를 수 있다는 가정이 가능하다. 백진환(2014)과 조현양, 백진환(2015)은 2번 성격 유형의 초등학교생을 대상으로 그들이 선호하는 도서와 그

8 한국도서관정보학회지(제48권 제2호)

들의 부족한 부분을 강화할 수 있는 내용을 포함하고 있는 도서를 분석하였다. 분석 결과 2번 유형의 초등학생들이 선호하는 도서, 즉 긍정적 특성을 강화할 수 있는 도서의 종류와 부정적 특성의 출현을 최소화하고 이를 보완할 수 있는 도서에 포함된 내용을 정리하였다. 정리한 내용은 다음의 <표 3>과 같다.

<표 2> 성격유형별 긍정적 특성과 부정적 특성

| 유형 | 긍정적 특성 | 부정적 특성 |
|----|------------------|-------------------------|
| 1 | 완벽함과 정의로움 | 독선적이고 완고하며 편협함 |
| 2 | 정이 많고 사람을 잘 돌봄 | 되돌려 받기를 원하며 감정에 치우치는 경향 |
| 3 | 목표 지향적이고 자신감에 넘침 | 자만하고 지나친 경쟁의식 |
| 4 | 자기 성찰과 표현력이 풍부 | 변덕스럽고 질투심이 강함 |
| 5 | 분석적이고 통찰력이 예리함 | 지적인 오만과 인색함 |
| 6 | 충실하고 남들에게 호감을 줌 | 지나치게 원칙에 집착하고 방어적 성향 |
| 7 | 상상력이 풍부하고 열정이 많음 | 산만하고 집중하지 못하고 즐거움에만 집착 |
| 8 | 단도직입적이고 권위가 있음 | 자기중심적이고 억지를 부림 |
| 9 | 포용력과 인내심이 강함 | 현실적인 대처를 못하고 수동적임 |

<표 3> 에니어그램 성격유형 2번을 위한 발전적 내용과 보완적 내용

| 발전적 내용 | 보완적 내용 |
|---|---|
| <ul style="list-style-type: none"> - 이타적, 자기희생적인 삶을 산 인물의 이야기 - 따뜻한 인간관계를 중심적으로 다루는 내용 - 가족적인 내용 | <ul style="list-style-type: none"> - 배려심이 깊어 자기주장이 내세우지 않으므로 리더십이 뛰어난 주인공이 나오는 내용의 도서 |
| <ul style="list-style-type: none"> - 친구와의 우정을 담고 있는 내용 - 가난한 국가의 아이들의 현실을 알려주는 내용 | <ul style="list-style-type: none"> - 착한아이 콤플렉스 등을 다루는 내용 |
| <ul style="list-style-type: none"> - 봉사, 나눔에 대한 내용 | <ul style="list-style-type: none"> - 자신의 자신감, 자존감을 높여줄 수 있는 이야기 |
| <ul style="list-style-type: none"> - 인물 간의 협동으로 성공적인 결과로 거두는 내용 - 외적 갈등을 잘 해소 해나가는 이야기 | <ul style="list-style-type: none"> - 타인을 먼저 생각하기보다는 본인의 역량 강화를 위한 사회교육 - 리더십이 나타나는 도서 |
| <ul style="list-style-type: none"> - 소시민적 이야기 - 동물/식물 기르기와 관련된 내용 | <ul style="list-style-type: none"> - 자신만의 세계가 독특했던 예술가, 위인에 관한 내용의 도서 |
| <ul style="list-style-type: none"> - 로맨스/드라마 등의 감정적인 내용 - 환상동화나 시적인 내용 및 장르 | <ul style="list-style-type: none"> - 특이하고 개성 넘치는 주인공이 나오는 내용의 도서 |
| <ul style="list-style-type: none"> - 2유형 특유의 협조적 역량 강화를 위한 자기계발서 - 인간관계/비즈니스/화술 | <ul style="list-style-type: none"> - 스스로의 내면을 인식하는데 도움이 되는 에세이 |

<표 1>에서 제시된 바와 같이 성격 유형을 대표하는 키워드의 추출이 가능하고 또한 그들이 선호하는 도서에 대한 분류가 가능하다면, 그리고 선호도서에 빈번하게 출현하는 어휘를 정확하게 판별할 수 있다면, 자동분류 기법을 활용하여 성격 유형별로 선호하는 도서에 대한

분류가 가능할 것이다. 예를 들어, 5번 성격 유형을 나타내는 사람의 특성은 사색가적인 기질을 가지고 있으며, 분석적이고, 주어진 일에 열중하면서도 나서지 않을 뿐만 아니라 지적 호기심이 많은 성격의 소유자로 표현된다. 이들의 특성을 대표하는 키워드는 지혜, 역사, 호기심, 선비, 지식, 성장, 자신감, 믿음 등으로 표현할 수 있다.

2. 성격 유형별 선호 도서

한윤옥과 조미아(2012)는 성격 유형을 대표하는 키워드와 해당 성격 유형의 어린이와 청소년들이 선호하는 도서에서 빈번하게 출현하는 키워드를 제시하였다. 상기에서 언급된 성격 유형별 특성으로부터 각각의 유형을 대표할 수 있는 키워드와 그들이 선호하는 도서에 빈번하게 출현하는 키워드를 비교한 결과는 다음의 <표 4>와 같다.

<표 4> 에니어그램 성격유형별 대표 키워드와 선호도서에 빈번하게 출현하는 키워드

| 성격 유형 | 키워드(성격설명) | 키워드(선호도서 추출) |
|-------|---|--|
| 1번 | 체계적, 객관적, 양심적, 원칙적, 진지함, 엄격함, 완벽함, 철저함 | 경제, 사회, 자연과학, 생태계, 운명, 성장, 고민, 철학, 성장, 전통문화, 역사, 인물, 그리움 |
| 2번 | 사교적, 친절, 친숙함, 대화, 친밀감, 사려 깊음, 타인 배려, 따뜻함, 관심 | 사회, 심리, 이미지, 자기관리, 추리, 탐정, 용기, 유머, 상상력, 우정, 성, 사랑, 극복, 투지, 감동, 재치 |
| 3번 | 자신감, 효율적, 좋은 인상, 성취감, 인정, 적응력, 자기개발, 체면 | 상식, 문화, 지혜, 교양, 정보, 요리, 인권, 민족, 여성, 입양, 꿈, 놀이 |
| 4번 | 창의적, 감정변화가 심함, 예민함, 독특한, 풍부한 감수성, 예술적 감성, 외로움 | 인간관계, 감동, 자기성장, 모험, 세계사, 한국사, 여왕, 예술, 우아함, 아름다움 |
| 5번 | 분석적, 사색적, 열중하는, 나서지 않는, 이성적, 정보수집, 지적 호기심 | 지리, 지혜, 역사, 호기심, 선비, 지식, 성장, 자신감, 어른, 학교, 믿음 |
| 6번 | 현신적, 가정적, 충실한, 회의적, 미리 준비하는, 책임감, 안정감 | 자아, 극복, 용기, 믿음, 칭찬, 성장, 소통, |
| 7번 | 긍정적, 열정적, 분위기를 이끄는, 다양한, 변화무쌍한, 사교적, 호기심, 모험심 | 자기개발, 일기, 상상력, 고난, 탐험, 사투, 모험, 순정, 무협, 공상, 아이디어 |
| 8번 | 독립적, 영웅적, 진취적, 다혈적, 강한 의지력, 지도력, 결단력, 자신감 | 경제, 가치관, 돈, 뉴스, 성적, 자기관리, 문화, 기후, 극복, 콤플렉스, 장애, 아동폭력, 생명, 전쟁, 학대, 화, 거짓, 정체성, 개혁, 환경문제 |
| 9번 | 편안한, 평화로운, 참가한, 무심한, 참을성이 많은, 온순한, 심적 여유 | 상처, 이해, 감동, 고민, 또래관계, 가족사랑, 유적, 자기에 |

(출처: 한윤옥, 조미아(2012) 수정)

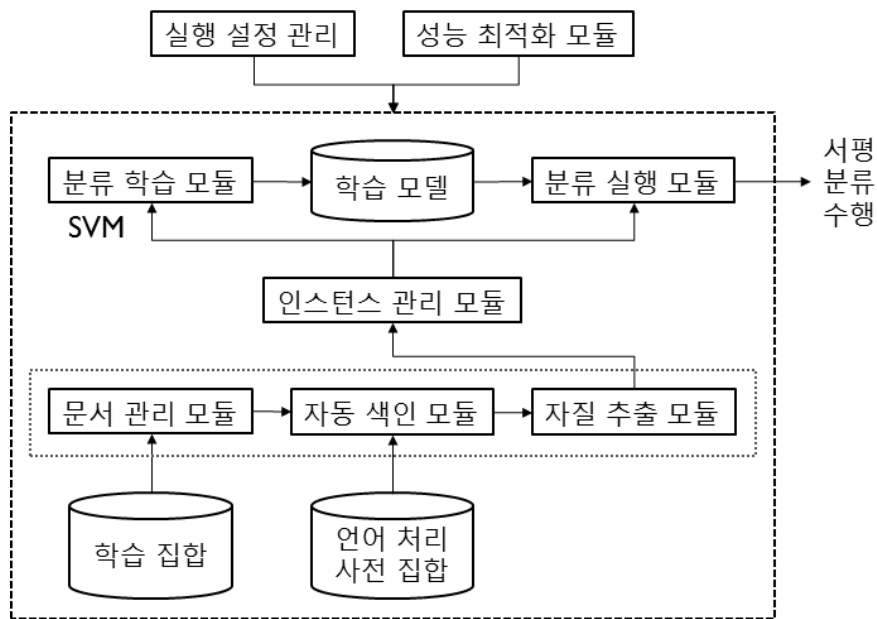
이 연구에서는 개인별 성향이나 성격 유형에 따라 선호하는 도서에 차이가 있음을 전제로, 국립어린이청소년도서관에서 제공하는 어린이·청소년을 위한 추천도서에 대한 책소개 정보를 활용하여 개별 책을 대표할 수 있는 키워드를 추출하고, 이를 성격 유형을 대표하는 키워

드와 비교함으로써 개인에게 적합한 도서를 합리적으로 추천할 수 있는 도서 자동분류시스템을 개발하였다.

Ⅲ. 도서 자동분류시스템 개발

1. 시스템 구성

이 연구에서 사용한 서평기반 도서 자동분류시스템은 서평데이터를 관리하기 위한 문서관리 모듈, 서평으로부터 키워드를 자동추출하기 위한 자동색인 모듈, 추출된 색인에 가중치를 부여하고 선별하기 위한 자질추출 모듈, 인스턴스 관리 모듈, 분류학습모듈, 그리고 성능 최적화 모듈로 구성되어 있으며, 그 구조는 다음의 <그림 1>과 같다.



<그림 1> 서평 자동분류 시스템 구조

2. 구성 요소별 기능

서평 자동분류시스템의 문서관리 모듈은 학습 집합으로 활용할 서평 데이터를 효율적으로 관리하고 탐색할 수 있도록 구성된 모듈이다. 서평레코드는 각 레코드의 ‘식별자’(#ID), ‘도

서명’(#TIT), 서평의 내용을 포함하고 있는 ‘서평 내용’(#CON), 그리고 성격유형 분류를 위한 ‘분류명’(#CAT)으로 구성된다. 입력문서의 구조는 다음의 <그림 2>와 같다.

```
@BOOK_VIEW
#ID=0
#TIT=디자이너
#CON=지혜와 용기, 그리고 끝없는 도전정신으로 꿈을 이룬 여성들과 그들의 다양한 직업 이야기를 들려주는 「꼬마이실의 여성 인물 이야기」 제16권 『빛나는 감각으로 세상을 설계한 디자이너』. 이 시리즈는 자신들의 실력과 의지로 세상을 깜짝 놀라게 했을 뿐 아니라, 역사를 바꾸어 놓기까지 한 열 명의 여성 이야기를 분야별로 담고 있다. 제16권에서는 창의력, 대담함, 노력, 결단력, 근성으로 명성을 얻은 남다른 업적을 남긴 세계적인 디자이너 열 명을 소개한다. 파격적인 아이디어를 생각해 내고 아름다운 형태와 선으로 표현했다는 공통점을 가진 이들의 이야기를 통해 자신만의 특별한 디자인을 직접 만들어보고 싶은 영감을 전해준다.
#CAT=3
```

<그림 2> 서평 레코드 구조

자동색인 모듈은 서평 레코드의 제목과 서평 내용에서 키워드를 자동으로 추출하는 모듈로 이 연구에서는 서울대학교 IDS(Intelligent Data Systems) 연구실에서 개발한 한국어 형태소분석시스템 ‘꼬꼬마’ 버전 2.0²⁾을 사용하였다. 꼬꼬마는 실험결과 분리되어야 할 단어를 단일 색인어로 추출하는 등의 몇 가지 문제점을 가지고 있음을 발견하였다. 예를 들어, “외로움만큼”과 같은 단어는 “외로움” + “만큼”으로 분리하여야 하지만 “외로움만큼”의 단일 색인어로 되는 문제점을 보였다. 또한 “민주항쟁·가출”와 같은 미등록어의 경우 “민주항쟁”과 “가출”의 독립된 두 명사로 추출되어야 함에도 불구하고 하나의 명사로 인식하는 등의 문제점을 내포하고 있었다. 자동색인 과정에서 발생하는 몇 가지 문제점에도 불구하고 대응량이 아닌 소량의 데이터를 가지고 실행한 이번 실험에는 약간의 수작업을 추가함으로써 작업의 수행이 가능하다.

자질추출 모듈은 자동색인모듈을 이용하여 개별 서평에서 추출된 색인어에 가중치를 부여하고 분류 성능에 중요한 색인어들을 선별하는 기능을 수행한다. 형태소 분석기를 통하여 1차 추출된 키워드에 가중치를 부여하기 위하여 사용한 자질은 서평에 나타나는 단어빈도(TF), 단어빈도의 로그값(LOG_TF), 역단어빈도(ITF), 역문헌빈도(IDF), 단어빈도와 역문헌빈도의 곱(TF_IDF), 그리고 용어의 발생 유무(0 또는 1)이다.

색인어 선택 방법은 문헌빈도(DF), 특정 단어에 대한 모든 분류와의 상호 정보(Mutual Information) 최대값(MI_MAX), 특정 단어에 대한 모든 분류와의 상호 정보 평균값(MI_AVG), 특정 단어에 대한 모든 분류와의 카이 제곱(Chi-square) 최대값(CHI_MAX), 그리고 특정 단어에 대한 모든 분류와의 카이 제곱 평균값(CHI_AVG)을 이용하였다.

2) 서울대학교 IDS 연구실에서 개발하여 공개한 한글형태소분석기로 <http://kkma.snu.ac.kr/documents/index.jsp>를 통하여 내려 받을 수 있음.

자질추출 모듈에서의 자질추출 과정은 우선 특정 문헌에 대한 색인어를 추출하여 색인어에 대한 가중치를 계산하고, 다음으로 색인어 집합에 대해서 색인어 선택 기법에 의한 색인어 필터링을 수행하게 된다. 이 때 지정된 임계치 이하로 계산되는 색인어들은 모두 배제하였다.

인스턴스관리 모듈은 추출된 자질을 바탕으로 개별 문헌을 학습에 적절한 형태의 자질 벡터 집합으로 변환하여 관리하는 모듈이다. 분류 학습/실행 모듈은 분류를 위한 기계학습 모듈로 본 연구에서는 LibSVM(Chang & Lin, 2011)과 LIBLINEAR(Fan, et al., 2008)의 라이브러리를 활용하였다. 성능최적화 모듈은 개별 문헌을 학습된 자질에 기초하여 문헌을 자동으로 분류함에 있어서 성능을 극대화하기 위한 모듈로 이 연구에서는 높은 성능을 나타내는 매개변수 조합을 찾기 위하여 10겹 교차검증을 활용하여 최적화를 수행하였다.

실행 설정관리 모듈은 다양한 설정을 통해서 자동분류시스템의 활용성을 극대화시킬 수 있도록 구성하였다. <그림 3>은 설정관리 모듈의 입력으로 들어가는 실제 자동 분류 설정 파일의 예시를 보여주고 있다.

```
# 분류 개수 지정.
category.number=9
# 문서집합 내에서 레코드를 시작하는 헤더 지정.
document.record.header=@BOOK_VIEW
# 원본 문서집합이 저장된 디렉토리 (다른 문서는 포함되어서는 안됨.)
document.directory=./bookcat/data
# 문서에 대한 색인 결과 저장 파일 지정.
document.info.file=./bookcat/bookcat.di
index.engine.mode=KOKOMA
# 색인 및 자질 추출이 완료된 역파일 저장 파일명 지정(가장 중요)
index.file=./bookcat/bookcat.index
# 기계 학습 엔진 선택
# LIBSVM(다양한 커널 적용 가능), LIBLINEAR(선형 커널만 가능한 대신 속도가 빠름)
train.engine.mode=LIBSVM
# 기계 학습 엔진에 입력되는 매개변수 지정
train.engine.param=-b 1 -c 8 -g 0.0625
# 학습 인스턴스 집합 저장 파일명 지정.
train.instance.file=./bookcat/bookcat.train
# 기계 학습이 완료된 모델 파일명 지정.
train.model.file=./bookcat/bookcat.model
# 학습이 완료된 모델을 기반으로 테스트를 수행할 실험 인스턴스 집합 파일명 지정.
test.instance.file=./bookcat/bookcat.train
# 문헌에서 추출된 자질 정보 저장 파일.
feature.file=./bookcat/bookcat.feature
```

<그림 3> 실행 설정 파일 예시 (일부)

위의 <그림 3>에서 보는 바와 같이, 설정 파일에는 시스템의 입력으로 활용되는 데이터 레코드의 양식은 물론 대상 데이터 위치 정보, 활용되는 기계학습 라이브러리, 학습 매개변수 및 최

종 결과 파일명 등을 지정할 수 있다. 이러한 세부적인 설정 정보를 활용하여 분류 성능에 대한 최적화를 진행할 수 있으며, 다양한 종류의 데이터를 바탕으로 자동 분류 실험이 가능하다.

IV. 데이터 실험 및 분석

1. 실험 데이터 및 실험 절차

이 연구에서 사용한 데이터는 국립어린이청소년도서관에서 유아, 초등저학년, 초등고학년, 그리고 청소년의 4등급으로 구분하여 제공하는 695권의 '사서추천도서' 가운데 501권의 유아 및 아동도서를 대상으로 하였다. 사서추천도서에 포함된 내용은 추천일시, 책소개(책에 대한 요약정보), 주제 구분, 그 외 간략한 서지정보 등이다. 이 중 '책소개'는 사서들에 의하여 작성된 책에 대한 내용 소개 및 요약 정보를 포함하고 있어서 해당 도서에 대한 개괄적인 내용 파악이 가능하여 성격유형별로 적합한 도서를 구분하는 데 사용이 가능하다. 책소개 정보를 이용하여 성격유형별로 적합한 도서를 분류하는 과정은 다음과 같다.

첫째, 유아 및 아동을 위한 도서에 포함된 책소개 정보를 활용하여 우선 성격유형별 적합도서를 3인의 에니어그램 성격 유형에 관한 내용을 학습한 독서전문가가 1차 수작업으로 분류하였다.

둘째, 분류된 도서를 최종적으로 에니어그램과 성격유형별 독서 분야의 전문가가 수정하고, 검증하는 단계를 거쳤다.

셋째, 전문가에 의하여 분류된 성격유형별 적합도서의 책소개 정보에 나타난 키워드를 이 연구를 통하여 개발한 서평자동분류시스템을 이용하여 추출하였다.

넷째, 국내에서 출판된 도서 가운데 현재 서점에서 판매중인 약 20만권을 대상으로 저자나 출판사에서 제공하는 서평을 이용하여 도서의 내용을 대표할 수 있는 키워드를 추출하고, 출현빈도와 출현한 키워드 간의 유사도를 측정하여 순위화하였다. 키워드를 추출하는 과정에서 특정 주제분야의 전문서적이 아닌 일반 도서를 대상으로 하는 것이기 때문에 추출된 용어가 책의 내용을 어느 정도 반영할 수 있는지를 확인하였고, 추출된 키워드를 기반으로 성격유형별 자동 분류를 위한 실험용 학습 집합을 구축하였다.

실험용 학습 집합을 이용하여 유아 및 아동용 도서 501권을 대상으로 에니어그램 성격 유형별로 자동 분류를 시행하였다. 다음의 <표 5>는 이 연구에서 활용한 도서 서평 자료의 성격 유형 분류별 데이터 건수를 보여준다. 표에서 보는 바와 같이 7번 유형에 해당하는 데이터 건수가 26건으로 가장 적고, 9번 유형에 맞는 데이터는 79건으로 가장 많은 것으로 나타났

다. 분류별 데이터 건수의 차이는 분류 성능과 직접적인 연관성이 있으므로 이를 감안한 성능 최적화 실험을 시도하였다.

〈표 5〉 성격 유형별 도서 건수

| 에니어그램 기반 성격 유형 | 데이터 건수 (도서) |
|----------------|-------------|
| 1번 유형 | 57 |
| 2번 유형 | 56 |
| 3번 유형 | 70 |
| 4번 유형 | 48 |
| 5번 유형 | 43 |
| 6번 유형 | 44 |
| 7번 유형 | 26 |
| 8번 유형 | 78 |
| 9번 유형 | 79 |
| 합 계 | 501 |

```
@BOOK_VIEW
#ID=372010
#ISBN=9788955821505
#AUT=오호선
#DAT=20110307120061115
#PUB=길벗어린이
#TIT=호랭이 꼬랭이 말놀이
#CON=아이들이 열광하는 본격 말놀이 책. 아이들은 말의 일관성이나 통일성이 아니라 무의미함과 허풍, 소란스러움을 즐기면서 상상력을 비약적으로 확장시킵니다. 그래서 모든 아이들은 말놀이를 열광적으로 좋아합니다. 자기가 아는 말을 시험하고 새로운 말을 지어내거나 엉뚱하게 뒤집으면서 즐거워합니다. 말의 의미보다는 리듬을 즐기고, 문답식으로 말을 주고받으며, 말 꼬리에 꼬리를 물고 이어지는 구성이 두드러지는 말놀이는 아이들이 말을 쉽게 익히도록 도와줄 뿐 아니라 상상력을 북돋아 주는 자양분이 됩니다. 서양에서는 전래동요에서 유래한 말놀이들이 오랫동안 어린이들의 사랑을 받아 왔습니다. 우리나라에서는 본격적인 말놀이라고 할 만한 이야기는 거의 찾아보기 어려웠습니다. 의성어, 의태어나 말을 반복하는 데서 오는 재미를 살린 간단한 이야기를 실은 책들이 있지만 대부분 상상력의 비약이나 난센스적인 요소가 없이 어린 아이들이 말을 쉽게 배우도록 돕는 기능을 주로 해 왔습니다. 이 책은 우리의 전래동요나 옛날이야기에 담긴 독특한 상상의 세계와 난센스를 살려 내면서 이제껏 주목되지 않은 짧은 이야기들을 찾아 말놀이를 공들여 다듬은 것입니다. 그동안 단편적으로만 소개되어 온 우리의 말놀이들을 적극적으로 발굴하고 재창작하여 우리 옛날이야기에 뿌리를 두면서도 오늘날 어린이들의 감각에 맞는 본격적인 말놀이로 펴냈다는 것에 이 책의 의의가 있습니다. 옛날이야기의 세계를 말놀이를 즐긴다. 이 책에는 옛날이야기에서 소재를 구해 새로 쓴 말놀이 15편이 담겨 있습니다. '꼬부랑'이란 말을 자꾸자꾸 붙이는 재미가 있는 <꼬부랑 할머니>, 꼬리에 꼬리를 물고 말이 이어지는 <불쌍한 두꺼비>와 <아까운 이야기>처럼 익숙한 형식의 이야기를 비롯해서 사람 머리에 사는 이가 이야기를 엮은 모습에서 나온 <아기 이와 엄마 이>, 해지는 소리가 뱅뮈는 <해님 달님>처럼 엉뚱하고 말도 안 되는 이야기들이 상상력을 자극합니다. <노래하는 장승>, <날씬한 >뽕다귀>, <딸랑귀신>은 줄거리가 있는 짧은 옛날이야기를 말의 재미와 리듬을 살린 신나는 말놀이를 새로 쓴 것입니다. 현대적 표현과 유머가 있는 그림 말이 안 되는 이야기를 듣다 보면 말도 안 되는 재미난 그림이 떠오릅니다. 그에 걸맞게 전통적인 소재에 현대적인 표현과 유머 감각을 조화시켜 말놀이의 독특한 세계를 잘 표현한 그림이 독자들에게 발랄한 웃음을 선사해 줍니다.
#CAT=9
```

〈그림 4〉 구축 데이터 예시 - 성격유형 9번에 적합한 도서

〈그림 4〉는 실험을 위하여 구축한 데이터의 예시이다. 각 레코드별 식별자와 함께 저자 및 ISBN이 포함되어 있고, 직접적인 분류 자질로 활용되는 도서 제목(TI)과 서평 데이터(CON)가 구축되어 있다.

2. 실험환경

이 연구에서는 책에 대한 소개나 서평정보를 활용하여 개인의 성격유형별로 적합한 도서를 추천하기 위하여 서평으로부터 자질을 추출하고 이에 기반하여 문헌을 자동으로 분류할 수 있는 자동 분류시스템을 개발하였다. 개발된 시스템에서 채택한 분류 모델은 선형 분류를 제공하는 LIBLINEAR 모델과 비선형 분류를 지원하는 LibSVM(RBF 커널) 모델을 활용하였다.

실험에서는 분류 자질 추출에 필요한 다양한 색인 가중치 부여 방법 및 자질 선택 방법을 적용하여 최적의 성능을 보이는 매개변수를 선택하고자 하였다. 이를 위해서 501건의 데이터에 대해서 기계학습 모델, 색인어 가중치 계산방법, 자질 선택 방법, 그리고 자질 선정 임계치 등 총 4가지 매개변수들을 변화시켜 가며 각 매개변수 설정 모드에 대해서 10겹 교차 검증을 수행하고 F-스코어를 계산하였다.

3. 성능 측정 결과 및 분석

실험에 활용된 2가지 기계학습 모델(비선형 커널 및 선형 커널) 각각에 대해서 총 6가지의 색인어 가중치 계산 방법과 자질 선택 방법, 그리고 10가지의 자질 선정 임계치 조합으로 구성된 360 개의 분류 모델들을 구성하고 각각의 성능을 측정하였다. 그 중 가장 높은 성능을 나타내는 분류 모델의 성능 측정 결과는 다음의 <표 6>과 같다.

<표 6> 개별 기계학습 모델별 최고 성능 측정 결과

| 학습 모델 | 색인어 가중치 계산 방법 | 자질 선택 방법 | 자질 선정 임계치 | 자질 선정 후 학습 집합 수 | F-스코어 |
|-----------------|---------------|----------|-----------|-----------------|--------|
| LibSVM (RBF 커널) | 역문헌빈도 | CHL_MAX | 0.3 | 474/501 | 0.5774 |
| LIBLINEAR | TF-IDF | MI_MAX | 0.6 | 499/501 | 0.6186 |

전체적으로는 선형 커널을 이용한 SVM(Support Vector Machine) 기반 학습 방법 (LIBLINEAR)이 비선형 분류를 지원하는 LibSVM(RBF 커널) 모델보다 더 나은 성능을 보여주고 있다. 비선형 커널을 활용한 모델은 전통적인 역문헌빈도 기반 색인어 가중치 계산 방법이 가장 높은 성능을 보이고 있으며, 선형 커널을 사용한 모델은 TF-IDF 가중치를 적용한 경우 0.62를 나타내어 최고 성능을 보이고 있는 것으로 실험 결과 나타났다. 추출된 자질들 중에서 유효한 자질을 선정하는 자질 선택 방법에는 문헌 분류에 많이 활용되고 있는 카이제곱 점수의 최고값을 적용한 경우와 상호 정보 점수의 최고값을 적용한 모델이 높은 성능을

보이고 있다. 또한 자질 선택 기법에 의해서 총 501건의 초기 문헌 수가 줄어들음을 보여주고 있는데 이는 자질 선정 임계치에 미달하는 색인어들만 포함한 문헌들이 학습 및 검증 과정에서 제거되는 현상 때문이다.

성능 측정 결과가 일반적인 뉴스 기사나 논문을 대상으로 한 문헌 분류 성능에 비해서 낮은 것으로 나타났으나, 명확한 의미 전달을 위하여 키워드의 선정을 우선시 하는 뉴스 기사나 합리적인 분류 기준이 존재하는 주제 분류에 비해서 성격 유형 기반 분류는 그 난이도가 월등히 높다는 것을 감안할 때, 초기 실험 결과로서 의미가 있다.

다음의 <표 7>과 <표 8>은 자질 선택 임계치 변화에 따른 자질 개수 및 성능 변화 결과 값을 보여 주고 있다.

<표 7> 자질 선택 임계치 변화에 따른 성능 변화 - LibSVM (RBF커널)

| CHI_MAX | 어휘 자질 개수 | 성능 (F1) |
|---------|----------|---------------|
| 0.1 | 15,728 | 0.2305 |
| 0.2 | 15,465 | 0.4618 |
| 0.3 | 15,465 | 0.5774 |
| 0.4 | 14,782 | 0.5084 |
| 0.5 | 14,566 | 0.4301 |
| 0.6 | 14,566 | 0.3781 |
| 0.7 | 14,566 | 0.3903 |
| 0.8 | 14,566 | 0.3243 |
| 0.9 | 14,177 | 0.1767 |

<표 8> 자질 선택 임계치 변화에 따른 성능 변화 - LIBLINEAR

| CHI_MAX | 어휘 자질 개수 | 성능 (F1) |
|---------|----------|---------------|
| 0.1 | 31,467 | 0.2939 |
| 0.2 | 31,467 | 0.3874 |
| 0.3 | 31,467 | 0.4769 |
| 0.4 | 31,467 | 0.5311 |
| 0.5 | 31,467 | 0.5501 |
| 0.6 | 31,467 | 0.6186 |
| 0.7 | 31,467 | 0.4044 |
| 0.8 | 31,278 | 0.1965 |
| 0.9 | 30,521 | 0.1111 |

자질 선택 임계치 변화에 따른 자질 개수 및 성능 변화를 살펴 본 결과, 우선 임계치에 따라 성능이 극단적으로 변화함을 알 수 있다. 이는 분류 성능에 있어서 자질 선택 방법이 결정적

인 역할을 수행함을 의미한다. 또한 비선형 커널(LibSVM)과 비교하여 선형 커널(LIBLINEAR)의 성능이 전반적으로 높을 뿐 아니라 살아남은 자질의 개수도 2배 정도 많은 것으로 나타났다. 개별 문헌을 수치적으로 표현하는데 있어서 기반 요소가 되는 어휘 자질의 규모가 크다는 의미는 문헌에 대한 표현 범위가 광범위함을 나타낸다. 따라서 학습 데이터의 규모가 커지면 커질수록 분류 성능의 상승도도 급격하게 상승할 수 있다는 예측이 가능하다.

<표 9> 성격 유형별 최상위 빈도 색인어 리스트

| 1번 유형 | | 2번 유형 | | 3번 유형 | | 4번 유형 | | 5번 유형 | | 6번 유형 | | 7번 유형 | | 8번 유형 | | 9번 유형 | |
|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|
| 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 | 색인어 | 빈도 |
| 아이 | 202 | 아이 | 249 | 아이 | 253 | 아이 | 215 | 이야기 | 170 | 아이 | 179 | 아이 | 260 | 아이 | 295 | 이야기 | 307 |
| 이야기 | 145 | 사람 | 213 | 친구 | 171 | 이야기 | 128 | 아이 | 161 | 사람 | 166 | 이야기 | 79 | 이야기 | 230 | 아이 | 303 |
| 사람 | 99 | 이야기 | 191 | 사람 | 156 | 사람 | 102 | 사람 | 140 | 자신 | 116 | 학교 | 68 | 사람 | 219 | 엄마 | 244 |
| 마음 | 92 | 어린이 | 143 | 이야기 | 154 | 친구 | 99 | 생각 | 108 | 이야기 | 103 | 친구 | 60 | 세상 | 155 | 사람 | 215 |
| 생각 | 83 | 생각 | 121 | 우리 | 119 | 생각 | 92 | 세상 | 106 | 엄마 | 96 | 작가 | 55 | 생각 | 151 | 역사 | 169 |
| 어린이 | 75 | 마음 | 120 | 수학 | 118 | 아빠 | 85 | 도서관 | 71 | 생각 | 90 | 선생님 | 54 | 그림책 | 142 | 우리 | 155 |
| 우리 | 65 | 엄마 | 106 | 생각 | 103 | 작가 | 80 | 우리 | 70 | 마음 | 77 | 엄마 | 49 | 작가 | 139 | 세계 | 147 |
| 그림책 | 62 | 친구 | 102 | 자신 | 94 | 우리 | 76 | 작가 | 70 | 아빠 | 77 | 세상 | 44 | 우리 | 135 | 문화 | 135 |
| 자신 | 61 | 우리 | 96 | 과학 | 89 | 엄마 | 72 | 자신 | 64 | 친구 | 72 | 그림 | 43 | 마음 | 133 | 생각 | 130 |
| 아빠 | 56 | 작가 | 94 | 문제 | 85 | 수로 | 72 | 친구 | 54 | 거북이 | 65 | 우리 | 40 | 친구 | 114 | 그림 | 127 |

마지막으로 <표 9>는 실험에 사용된 데이터 집합에서 각 성격 유형별로 추출한 최상위 빈도 색인어 리스트를 보여주고 있다. <표 9>에서 나타난 바와 같이 유형별로 차이는 보이지만 대체적으로 유사한 색인어들이 자주 등장하고 있다. 이러한 현상은 2가지 측면에서 그 이유를 찾을 수 있다. 첫째, 이 연구에서 자동 분류를 위한 대상이 어린이를 위한 도서로 한정되어 있다. 따라서 도서에 출현하는 용어의 범위가 한정적일 뿐만 아니라 ‘어린이’, ‘이야기’, 그리고 ‘아이’와 같이 어린이들의 관심을 집중시킬 수 있는 용어들이 빈번하게 나타나는 것이 일반적이다. 이러한 측면에서 도서에서 공통적으로 빈번하게 출현하는 용어, 즉 도서의 내용에 대한 변별력이 떨어지는 용어 혹은 문헌분리도가 낮은 키워드는 추출시 가중치를 달리하는 것도 하나의 방법이 될 수 있을 것이다.

그 다음으로 성격 유형을 기준으로 도서를 더 정확하게 분류하기 위해서는 분류별 색인어 리스트를 활용한 단순 매칭 혹은 검색 기법으로는 한계가 있음을 의미한다. 즉, 키워드 기반의 서평 자동분류시스템을 이용한 자동 분류의 정확도를 높이기 위해서는 시스템 자체의 성

능도 중요하지만, 개인별 성격유형을 보다 명확하게 표현할 수 있는 색인어 집합의 지속적인 확대와 더불어 대용량의 학습 집합의 구축이 선행되어야 할 것으로 판단된다. 일반적으로 한정된 주제를 포함하고 있으며, 또한 이를 분류하기 위한 분류체계가 존재하는 분야별 전문도서와는 달리 독서의 대상이 되는 도서의 경우 주제보다는 장르나 형태에 따른 분류가 일반적이기 때문에 키워드 기반의 분류는 한계점을 가질 수 있다. 결론적으로 개인별 성향에 맞는 도서를 분류하거나 추천하기 위해서는 정교하게 구축된 대량의 학습 집합을 바탕으로 보다 높은 수준의 기계학습 기반 분류 모델이 필요하다.

V. 결론 및 제언

이 연구의 목적은 개인별 성향이나 성격 유형에 따라 선호하는 도서에 차이가 있음을 전제로, 국립어린이청소년도서관에서 제공하는 어린이·청소년을 위한 추천도서에 대한 책소개 정보를 활용하여 개별 책을 대표할 수 있는 키워드를 추출하고, 이를 성격 유형을 대표하는 키워드와 비교함으로써 개인에게 가장 적합한 도서를 합리적으로 추천할 수 있는 서평 자동분류시스템을 개발하는 것이다. 연구에서 사용한 데이터는 국립어린이청소년도서관에서 유아, 초등저학년, 초등고학년, 그리고 청소년의 4등급으로 구분하여 제공하는 695권의 '사서 추천도서' 가운데 501권의 유아 및 아동도서를 대상으로 하였으며, 그 결과는 다음과 같다.

첫째, 국립어린이청소년도서관에서 제공하는 유아 및 아동용을 위한 추천도서에 포함된 책소개 정보를 활용하여 우선 성격유형별 적합도서를 3인의 에니어그램 성격 유형을 학습한 도서 전문가가 1차 수작업으로 분류하고, 분류된 도서를 최종적으로 에니어그램과 성격유형별 도서 분야의 전문가가 검증하는 단계를 거쳤다.

둘째, 전문가에 의하여 분류된 성격유형별 적합도서의 책소개 정보에 나타난 키워드를 다양한 가중치 지정 및 자질 추출 기법을 지원하는 서평자동분류시스템을 개발하였으며, 개발된 시스템에서 채택한 분류 모델은 선형 분류를 제공하는 LIBLINEAR 모델과 비선형 분류를 지원하는 LibSVM(RBF 커널) 모델을 활용하였다.

셋째, 전문가 집단에 의하여 분류된 성격유형별 적합도서의 책소개 정보에 나타난 키워드를 이 연구에서 개발한 서평 자동분류시스템을 이용하여 자동분류를 실험적으로 수행하였으며, 이를 기반으로 도서 분류를 위한 학습 집합을 구축하였다.

넷째, 도서로부터 추출하여 군집화한 키워드와 에니어그램 성격유형별로 대표하는 키워드를 매치시켜 성격유형별로 적합한 도서를 추천하였다.

실험에 활용된 2가지 기계학습 모델(비선형 커널 및 선형 커널) 각각에 대해서 총 6가지의

색인어 가중치 계산 방법과 자질 선택 방법, 그리고 10가지의 자질 선정 임계치 조합으로 구성된 360 개의 분류 모델들을 구성하고 각각의 성능을 측정하였다. 전체적으로는 선형 커널을 이용한 SVM 기반 학습 방법(LIBLINEAR)이 비선형 분류를 지원하는 LibSVM(RBF 커널) 모델보다 더 나은 성능을 보이는 것으로 나타났다. 다만 성능 측정 결과는 뉴스 기사나 논문을 대상으로 한 문헌 분류 성능에 비해서 낮은 것으로 나타났으나, 합리적인 분류 기준이 존재하는 뉴스기사나 주제 분류에 비해서 성격 유형 기반 분류는 그 난이도가 높다는 것을 감안할 때, 초기 실험 결과로서의 의미는 부여할 수 있다.

성격 유형을 기준으로 도서를 분류하기 위해서는 분류별 색인어 리스트를 활용한 단순 매칭 혹은 검색 기법으로는 분명히 한계가 존재한다. 즉, 키워드 기반의 서평자동분류시스템을 이용한 자동 분류의 정확도를 높이기 위해서는 시스템 자체의 성능도 중요하지만, 개인별 성격유형을 보다 명확하게 표현할 수 있는 색인어 집합의 지속적인 확대와 더불어 대용량의 학습 집합의 구축이 선행되어야 할 것이다. 또한 개인별 성향에 맞는 도서를 분류하거나 추천하기 위해서는 정교하게 구축된 대량의 학습 집합을 바탕으로 보다 높은 수준의 기계학습 기반 분류 모델이 필요하다.

이 연구를 통하여 개발된 서평 자동 분류시스템은 지금까지 연구된 상황이나 성격 유형에 따른 도서 목록 생산 분야의 도서를 보다 신속하고 정확하게 분류할 수 있는 기초자료로 활용이 가능할 뿐만 아니라 양적 독서에서 탈피하여 질적 독서에 대한 가능성을 높이기 위한 수단으로 개인별 성향이나 상황에 적합한 도서를 추천할 수 있을 것이다. 또한 독자에게는 읽을 거리에 보다 용이하게 접근할 수 있는 채널의 제공할 수 있으며, 출판사나 저작자들에게는 책에 대한 홍보활동 자료로 활용이 가능하다.

참고문헌

- 김경. 2016. 지적장애학생 성격유형에 따른 미술치료프로그램 개발. 『한국수산해양교육』, 28(1): 266-278.
- 김국선. 2013. 어린이 성격유형별 학습능력 향상을 위한 공간디자인 구축 방안: 에니어그램 성격 특성 분석을 통하여. 『한국가구학회지』, 24(1): 42-50.
- 김기열. 2015. 중학생의 성격유형에 따른 기술적 사고성향 분석. 『한국실과교육학회지』, 28(2): 135-152.
- 김수아, 오성탁, 이지형. 2015a. 범주별 고유정보를 고려한 블로그 포스트의 자동 분류. 『한국컴퓨터정보학회 학술발표논문집』, 23(1): 11-14.

- 김수아, 조희선, 이현아. 2015b. 다양한 어휘 가중치를 이용한 블로그 포스트의 자동 분류. 『한국마린엔지니어링학회지』, 39(1): 58-62.
- 김관준, 이재운. 2012. 디스크립터 자동 할당을 위한 저자키워드의 재분류에 관한 실험적 연구. 『정보관리학회지』, 29(2): 225-248.
- 모슬기, 김영숙. 2009. 중·고등학교 영어 학습자들의 성격 유형에 따른 영어 독해전략 연구. 『영어어문교육』, 15(4): 291-311.
- 백진환, 한윤옥. 2014. 에니어그램 성격유형을 적용한 독서지도의 효과 연구. 『한국문헌정보학회지』, 48(2): 45-64.
- 윤은성. 2001. 『한국형 에니어그램 프로파일』, 서울: 한국에니어그램연구소.
- 이병열. 2015. 여행업 종사자의 성격유형과 업무성과의 관계 연구. 『관광·레저연구』, 27(2): 387-406.
- 장재영, 김정민, 이신영. 2012. 단어 패턴 빈도를 이용한 한국어 영화평 자동 분류기법. 『한국정보과학회 학술발표논문집』, 39(1C): 51-53.
- 장지숙, 이해영. 2009. 맥락정보를 이용한 기록 자동분류시스템 설계. 『한국기록관리학회지』, 9(1): 151-73.
- 정덕현. 2015. 성격유형별 인지적 사고 및 글쓰기 과정 연구. 『교양교육연구』, 9(3): 435-471.
- 조현양, 한윤옥, 백진환. 2015. 성장지향점 도서를 활용한 독서지도 효과에 관한 연구: 에니어그램 성격 2번 유형을 중심으로. 『한국문헌정보학회지』, 49(1): 69-88.
- 최형림 외. 2003. 기계학습 기법을 이용한 전자계시관 질문 자동 분류. 『한국지능정보시스템학회 학술대회논문집』, 2003. 5: 313-318.
- 한윤옥. 2008. 『독서교육과 자료의 활용』. 서울: 한국도서관협회.
- 한윤옥, 조미아. 2011. 독서교육에 있어서 청소년의 성격유형과 독서성향의 관련성 연구. 『한국도서관·정보학회지』, 42(4): 69-88.
- 한윤옥, 조미아. 2012. 『어린이·청소년의 성격유형별 선호도서 목록개발 및 독서지도 방안 연구』, 서울: 국립어린이청소년도서관.
- 한윤옥 외. 2012. 성격유형을 적용한 청소년 독서지도 프로그램 운영 방안에 관한 연구. 『한국문헌정보학회지』, 46(1): 77-97.
- 한윤옥 외. 2012. “어린이의 성격유형에 따른 선호도서의 특성에 관한 연구.” 『한국문헌정보학회지』, 46(3): 5-34.
- Chang, Chih-Chung and Lin, Chih-Jen. 2011. “LIBSVM: A Library for Support Vector Machines.” *ACM Transactions on Intelligent Systems and Technology*, 2(3): 1-27.
- Fan, R.E. et al. 2008. “LIBLINEAR: A Library for Large Linear Classification.” *Journal*

of Machine Learning Research. 9, 1871–1874.

Russell, Anne L. 2002. “MBTI Personality Preferences and Diverse Online Learning Experiences.” *School Libraries Worldwide*, 8(1): 25–40.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Chang, Jae-Young, Kim, JungMin, and Lee, SinYoung. 2012. “Automatic Classification of Korean Movie Reviews Using a Word Pattern Frequency.” *Proceedings of the Korean Information Science Society Conference*. 39(1C): 51–53.
- Cho, Hyun-Yang, Han, Yoon-Ok, and Paek, Jin-Hwan. 2015. “A Study on the Effects of reading Guidance Using a Happy Point Books, Focused on Enneagram Personality Type 2.” *Journal of the Korean Society For Library and Information Science*, 49(1): 69–88.
- Choi, Hyung-Rim et al. 2003. “An Automatic Question Routing System Using Machine Learning.” *Proceedings of the Korean Intelligent Information Systems Society*. May 2003: 313–318.
- Han, Yoon-Ok, & Cho, Miah. 2011. “A Study on the Relations of Personality Type and Reading Tendency of the Youth in Reading Education.” *Journal of Korean Library and Information Science Society*, 42(4): 69–88.
- Han, Yoon-Ok, & Cho, Miah. 2012. *A Study on the Development of Preference Book List and the Operation of the Reading Guidance Program, Based on Children and Adolescents’ Personality Type*. Seoul: National Library for Children & Young Adults.
- Han, Yoon-Ok, et al. 2012. “A Study on the Characteristics of Preference Books According to the Children’s Personality Type.” *Journal of the Korean Society For Library and Information Science*., 46(3): 5–34.
- Han, Yoon-Ok, et al. 2012. “A Study on the Operation of the Reading Guidance program for the Youth According to Personality Type.” *Journal of the Korean Society For Library and Information Science*, 46(1): 77–97.
- Han, Yoon-Ok. 2008. *Education and the Use of Reading Material*. Seoul : Korean Library Association.
- Jang, Ji-Sook, and Rieh, Hae-Young. 2009. “Design of Automatic Records Classification

- System Using Contextual Information.” *Journal of Korean Society of Archives and Records Management*. 9(1): 151–173.
- Jeong, Dukhyun. 2015. “Research on Cognitive Thinking and Writing Process as a Personality Type.” *Korean Journal of General Education*. 9(3): 435–471.
- Kim, Ki–Yeol. 2015. “Analysing Technological Thinking Disposition by Students’ Personality Types in Middle School.” *Journal of Korean Practical Arts Education*. 28(2): 135–152.
- Kim, Kook–Sun. 2013. “Space Design for Enhancing Learning Ability with Children’s Character Type: Through Analyzed Enneagram Tool.” *Journal of the Korea Furniture Society*. 24(1): 42–50.
- Kim, Kyong. 2016. “Development of Art Therapy Program According to the Personality Type of Student with Intellectual Disabilities.” *The Korean Society Fisheries and Science Education*. 28(1): 266–278.
- Kim, Pan Jun, and Lee, Jae Yun. “A Study on the Reclassification of Author Keywords for Automatic Assignment of Descriptors.” *Journal of the Korean Society for Information Management*. 29(2): 225–248.
- Kim, Suah, Jho, Hee–Sun, and Lee Hyun Ah. 2015. “Automatic Classification of Blog Posts Considering Category–specific Information.” *Journal of the Korean Society of Marine Engineering*, 39(1): 58–62.
- Kim, Suah, Oh, Sungtak, and Lee, Jee–Hyong. 2015. “Automatic Classification of Blog Posts Considering Category–specific Information.” *Proceedings of the Korean Society of Computer Information Conference*, 23(1): 11–14.
- Lee, Byeong–Yeul. 2015. “A Study on the Relationship of Personality Type and Job Performance of Employees’s on travel Industry.” *Journal of Tourism and Leisure Research*. 27(2): 387–406.
- Mo, Seul–ki and Kim, Youngsook. 2009. “A Study on the English Reading Strategies of Korean Secondary EFL Learners by Their Personality Type.” *English Language & Literature Teaching*. 15(4): 291–311.
- Paek, Jin–Hwan & Han, Yoon–Ok. 2014. “A Study on Effect of Reading Guidance Program based on Enneagram of Personality” *Journal of the Korean Society For Library and Information Science*, 48(2): 45–64.
- Yun, U. 2001. *Korean Enneagram Profile*. Seoul: Korea Enneagram Education Foundation.