

# 준지도학습 방법을 이용한 한국어 서답형 문항 자동채점 시스템

천민아<sup>○</sup>, 서형원, 김재훈, 노은희<sup>†</sup>, 성경희<sup>†</sup>, 임은영<sup>†</sup>  
한국해양대학교, 컴퓨터공학과  
한국교육과정평가원<sup>†</sup>

minah014@outlook.com, wonn24@gmail.com, jhoon@kmou.ac.kr  
{noro, kelly9147, elim}@kice.re.kr<sup>†</sup>

## Korean Automated Scoring System for Supply-Type Items using Semi-Supervised Learning

Min-Ah Cheon<sup>○</sup>, Hyeong-Won Seo, Jae-Hoon Kim,  
Eun-Hee Noh<sup>†</sup>, Kyung-Hee Sung<sup>†</sup>, EunYoung Lim<sup>†</sup>

Department of Computer Engineering, Korea Maritime and Ocean University  
Korea Institute for Curriculum and Evaluation<sup>†</sup>

### 요 약

서답형 문항은 학생들의 종합적인 사고능력을 판단하는데 매우 유용하지만 채점할 때, 시간과 비용이 매우 많이 소요되고 채점자의 공정성을 확보해야 하는 어려움이 있다. 이러한 문제를 개선하기 위해 본 논문에서는 서답형 문항에 대한 자동채점 시스템을 제안한다. 본 논문에서 제안하는 시스템은 크게 언어 처리 단계와 채점 단계로 나뉜다. 첫 번째로 언어 처리 단계에서는 형태소 분석과 같은 한국어 정보처리 시스템을 이용하여 학생들의 답안을 분석한다. 두 번째로 채점 단계를 진행하는데 이 단계는 아래와 같은 순서로 진행된다. 1) 첫 번째 단계에서 분석 결과가 완전히 일치하는 답안들을 하나의 유형으로 간주하여 각 유형에 속한 답안의 빈도수가 높은 순서대로 정렬하여 인간 채점자가 고빈도 학생 답안을 수동으로 채점한다. 2) 현재까지 채점된 결과와 모범답안을 학습말뭉치로 간주하여 자질 추출 및 자질 가중치 학습을 수행한다. 3) 2)의 학습 결과를 토대로 미채점 답안들을 군집화하여 분류한다. 4) 분류된 결과 중에서 신뢰성이 높은 채점 답안에 대해서 인간 채점자가 확인하고 학습말뭉치에 추가한다. 5) 이와 같은 방법으로 미채점 답안이 존재하지 않을 때까지 반복한다. 제안된 시스템을 평가하기 위해서 2013년 학업성취도 평가의 사회(중3) 및 국어(고2) 과목의 서답형 문항을 사용하였다. 각 과목에서 1000개의 학생 답안을 추출하여 채점시간과 정확률을 평가하였다. 채점시간을 전체적으로 약 80% 이상 줄일 수 있었고 채점 정확률은 사회 및 국어 과목에 대해 각각 98.7%와 97.2%로 나타났다. 앞으로 자동 채점 시스템의 성능을 개선하고 인간 채점자의 집중도를 높일 수 있도록 인터페이스를 개선한다면 국가수준의 대단위 평가에 충분히 활용할 수 있을 것으로 생각한다.

주제어: 자동채점, 기계학습 기반 분류기, 준지도 학습

### 1. 서론

대학수능능력시험과 같이 대규모로 시행되는 시험의 경우, 채점에 드는 비용과 시간을 줄이고 채점상의 오류를 줄이기 위해 시험 문항을 선택형(객관식) 문항으로 제한하고 있다. 그러나 선택형 문항만으로는 학생들의 창의력, 문제해결능력 등의 고차원적 사고력을 평가할 수 없다는 문제점이 있다[1-4]. 이런 문제점을 해결하기 위해, 평가문항 유형을 다양화하여 종합적인 이해와 문제해결 능력을 측정할 수 있도록 노력하고 있다[1-2]. 학생들의 종합적인 사고력을 판단하는데 유용한 문항으로 단답형, 서술형, 논술형(논문형) 등이 있으며 이런 문항들을 서답형 문항이라고 부른다[2-3]. 그러나 서답형 문항은 채점에 드는 비용과 시간의 소요가 크며, 채점을 위한 전문 인력의 확보가 어렵다는 점과 채점자마다 채점 기준이 다를 수 있어 채점의 공정성과 채점 결과의 신뢰성을 확보하는데 어렵다는 단점이 있다[2].

이와 같은 문제를 해결하기 위해 국외에서는 자동 채

점 시스템[2, 4]이 자주 활용하고 있다.

본 논문에서는 한국어 서답형 문항의 자동 채점 시스템을 제안한다. 제안된 시스템은 크게 언어 처리 단계와 채점 단계로 나뉜다. 언어 처리 단계에서는 형태소 분석과 같은 한국어 정보처리 시스템을 이용하여 학생들의 답안을 분석한다. 이 단계는 문서 정규화, 형태소분석 및 품사부착, 구뭉음, 바꿔쓰기, 의존구조분석으로 구성된다. 언어 처리 단계가 끝나면 채점 단계를 진행하는데 채점 단계는 크게 다섯 단계로 진행된다. 첫 번째 단계는 초기 채점 단계로서 언어 처리 결과가 완전히 일치하는 답안들을 하나의 유형으로 간주하여 해당 유형에 속한 답안의 빈도순으로 정렬한다. 그 뒤 인간채점자가 고빈도 학생 답안을 수동으로 채점한다. 두 번째 단계는 현재까지 채점된 결과와 모범답안을 학습말뭉치로 간주하여 자질을 추출하고, 선택한 자질의 가중치를 학습한다. 세 번째 단계는 미채점 답안의 군집화로 학습된 결과를 바탕으로 미채점 답안을 군집화하여 분류한다. 네 번째 단계는 검증에 위한 과정으로서 분류된 결과 중에서 신뢰성이 높은 채점 답안에 대해서 인간 채점자가 확

인하고 학습말뭉치에 추가한다. 다섯 번째 단계는 두 번째 단계부터 네 번째 단계를 미채점 답안이 존재하지 않을 때까지 반복한다. 이와 같은 일련의 과정을 효과적으로 수행할 수 있고 인간 채점자가 채점에 집중할 수 있도록 사용자 인터페이스를 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 서답형 문항 자동채점 시스템 개발을 위해 필요한 관련 연구들을 소개한다. 3장에서는 구현한 서답형 문항 자동채점 시스템에 대해 설명한다. 4장에서는 구현된 자동채점 시스템을 평가하고 결과를 분석한다. 마지막으로 5장에서는 결론 및 향후 연구에 대해서 논의한다.

## 2. 관련 연구

이 장에서는 해외의 자동채점 시스템에 대해 간략히 소개하고, 본 논문에서 제안하는 서답형 문항 자동채점 시스템에 이용할 기술들을 소개한다.

### 2.1 서답형 문항 자동채점 시스템

영어 자동 채점 시스템은 1960년대에 시작되어 널리 사용되고 있으며 PEG(Project Essay Grader), IEA(Intelligent Essay Assessor), e-Rater(electronic essay Rater)와 Criterion, IntelliMetric과 MY Access!, BETSY(Bayesian Essay Test Scoring sYstem), MarkIT, C-rater 등이 있다[5-6]. 많은 경우 기계학습 방법을 사용하고 있다.

### 2.2 준지도학습

지도학습(supervised learning) 방법은 정답이 부착된 데이터를 학습 자료로 사용하여 분류기(classifier)를 학습하고 학습되지 않은(정답이 부착되지 않은) 데이터를 입력 받아서 정답을 부여하는 기계학습 방법이다. 지도학습 방법은 학습 자료를 구축하기 위해서 전문가의 도움이 필요하고 이 과정에서 많은 시간과 노력이 소요된다. 한편 비지도학습(unsupervised learning) 방법은 지도학습 방법과 달리 정답이 부착되지 않은 자료를 사용하여 그 자료의 구조나 관계를 파악하여 패턴을 분류한다. 위의 두 가지 학습 방법을 결합한 기계학습 방법을 준지도학습(semi-supervised learning)이라고 한다. 준지도학습은 정답이 부착된 자료와 정답이 부착되지 않은 자료를 모두 학습 자료로 사용하는 방법으로, 학습이 진행될 때마다 정답이 부착되지 않은 자료에 대해 정답이 부착되고 이를 학습 자료로 사용함으로써 학습 자료의 양이 점점 늘어나서 분류기의 정확률 및 신뢰성을 지속적으로 개선하는 방법이다[7].

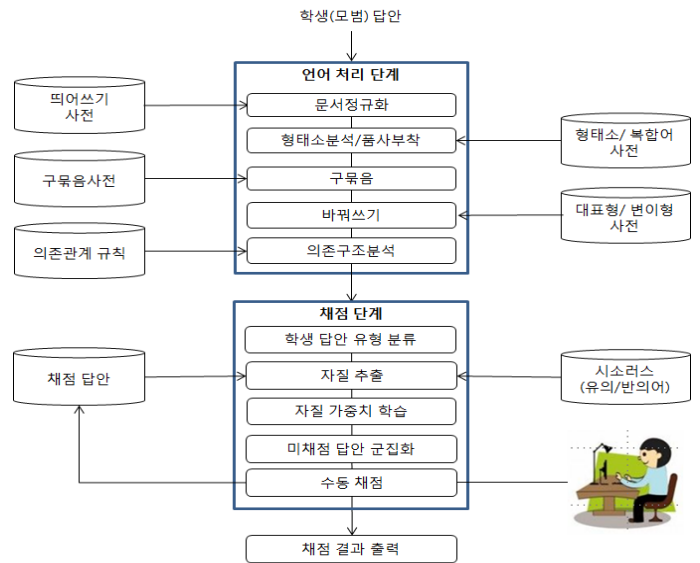
## 3. 준지도학습을 이용한 한국어 서답형 문항 자동채점 시스템

이 장에서는 제안된 서답형 문항의 자동채점 시스템을 기술한다. 먼저 전체 시스템의 구조에 대해서 간략히 소개하고 준지도학습을 이용한 자동채점 알고리즘

에 대해 자세하게 설명한다. 3.1절에서 서답형 문항 자동채점 시스템의 전체 시스템 구조에 대해서 설명하고, 3.2절과 3.3절에서는 각각 언어 처리 기능과 채점 기능에 자세히 기술한다.

### 3.1 전체 시스템 구조

[그림 1]은 서답형 문항 자동채점 시스템의 전체 시스템 구조도다. [그림 1]에서 나타난 것과 같이 서답형 답안 자동채점 시스템은 크게 ‘언어 처리 단계’와 ‘채점 단계’로 나뉜다.



[그림 1] 시스템의 전체 시스템 구조도

‘언어 처리 단계’는 학생 답안이나 모범 답안(문제 출제자가 작성한 예시 답안)에 대해 자동채점에 필요한 언어 정보를 분석하는 단계다.

‘채점 단계’는 학생 답안과 모범 답안의 언어 처리 결과를 군집화하여 각 군집에 점수를 부여하는 단계다. 초기 군집은 언어 처리 결과가 완전히 일치하는 문장들의 집합이다. 이 군집을 문장 빈도 역순으로 정렬하여 상위의 군집에 대해 인간 채점자가 채점을 수행한다. 이렇게 채점된 답안들과 모범 답안들을 학습 자료로 간주하여 자질을 추출하고 자동 채점 분류기를 학습한다. 학습된 시스템은 미채점 학생 답안을 자동으로 채점하고 채점 결과의 신뢰성이 답안에 대해서 인간 채점자가 확인한다. 이 과정은 미채점 답안이 없을 때까지 반복한다. 이 과정이 반복될수록 시스템의 신뢰성이 점점 높아지고 그 결과로 인간 채점자가 수동으로 채점해야 할 학생 답안 수는 점점 감소하게 된다.

### 3.2 언어 처리 단계

이 절에서는 본 논문에서 제안하는 시스템의 세부 기능 중 언어 처리 단계에 대해서 설명한다. 언어 처리 단계는 문서 정규화, 형태소 분석 및 품사 부착, 구문분석, 바꿔쓰기, 의존구조분석으로 구성되어 있으며, 서답형 문항에 따라서 인간 채점자가 선택적으로 적용할 수 있다.

### 1) 문서정규화(text normalization)

문서정규화는 문장부호 제거와 띄어쓰기 교정, 그리고 철자 교정이 가능하다. 문장부호 제거는 정규표현식(regular expression)을 통해 간단하게 구현하였다. 띄어쓰기(spacing)는 세종말뭉치[8]를 띄어쓰기 사전으로 사용하여 기계학습 방법(maximum entropy)을 통해 구현하였다. 철자교정(spelling correction)은 최소편집거리(minimum edit distance) 알고리즘을 이용하여 구현하였다.

### 2) 형태소 분석 및 품사 부착

형태소 분석(morphological analysis)은 CYK 알고리즘을 수정하여 사용하였다[9]. 이 알고리즘을 이용하면 어미의 활용이 존재하지 않을 경우, 사전 정보만으로도 가능한 모든 형태소 분석 결과를 찾아낼 수 있다는 장점이 있다. 그러나 한국어의 경우에는 “아름다운”과 같이 동사의 원형과 어미가 결합할 때(예: “아름답 + ㄴ”) 다양한 형태로 철자의 변형이 발생한다. 이 경우에는 CYK 알고리즘으로 처리할 수 없기 때문에 용언(동사와 형용사)에 대해서는 부분적인 기본식 사전을 활용하여 이 문제를 해결하였다.

품사 부착(part-of-speech tagging)은 형태소 분석 결과를 이용해서 격자 구조(lattice structure)를 구성하고 구성된 격자 위에 세종 말뭉치에서 구해진 문맥 확률( $\Pr(t_i|t_{i-1})$ , contextual probability)과 어휘 확률( $\Pr(w_i|t_i)$ , lexical probability)를 적재하여 가중치 네트워크(weighted network)를 형성한다. 가중치 네트워크에서 가장 적절한 경로(longest path)를 찾으면 그 결과에 품사 부착 결과가 된다.

### 3) 구 묶음(chunking)

구 묶음(chunking)은 띄어쓰기와 같이 기계학습 방법인 최대 엔트로피 모델(maximum entropy model)과 구 묶음 사전을 이용하여 구현하였다. 구 묶음을 수행하기 위해서는 반드시 형태소 분석 및 품사 부착이 선행되어야 한다.

### 4) 바꿔쓰기(paraphrasing)

바꿔쓰기(paraphrasing)는 단어 치환(word replacement)을 사용하여 구현하였다. 현재 이 기능은 조사를 대표조사로, 어미를 대표어미로, 단어를 동의어로 치환하도록 구현하였다. 대표형/변이형 사전(지식 베이스)에 해당 정보가 포함되어 있지 않으면 입력된 단어를 그대로 출력한다. 따라서 대표형/변이형 사전이 완전히 비어 있다면 입력과 출력은 똑같으며 이 사전은 문제의 영역이나 과목 등에 따라 다르므로 문제를 채점할 때 이점을 고려하여 대표형 사전을 보완하거나 수정할 필요가 있다. 바꿔쓰기를 하기 위해서는 반드시 형태소 분석 및 품사 부착이 선행되어야 한다.

### 5) 의존구조분석

의존구조분석은 의존관계 규칙을 기반으로 구문 분석

을 수행한다. 여기에서 의존관계 규칙이란 의존문법을 말하며, 의존소(dependent)와 지배소(governor)의 관계를 문법으로 표현한 것이다. 지배소는 의존관계에 있는 언어 요소들 중 의미의 중심이 되는 요소이며, 의존소는 지배소가 갖는 의미를 보완해주는 요소를 의미한다. 이 기능은 스웨덴의 Växjö 대학교와 Uppsala 대학교에서 공동으로 개발한 의존구문 분석기인 MaltParser를 사용하여 구하였다[10].

### 3.3 채점 단계

이 절에서는 채점 단계에 수행하는 기능들과 자동채점 알고리즘에 대해 설명한다. 채점 단계는 [그림 1]과 같이 학생 답안 유형 분류, 자질 추출, 자질 가중치 학습, 미채점 답안 군집화, 수동 채점으로 구성되어 있다.

#### 1) 학생 답안 유형 분류

언어 처리 결과가 완전히 일치하는 학생 답안들을 하나의 유형(군집)으로 간주하여 각 유형에 속한 답안 수가 높은 순으로 정렬하여 인간 채점자가 상위 유형에 대해서 수동으로 채점한다. 상위 몇 개의 유형을 채점할 것인지는 채점자가 임의로 결정한다. 이렇게 채점된 답안들은 모범 답안과 함께 초기 학습 자료로 사용된다.

#### 2) 자질 추출 및 가중치 학습

초기 학습 자료를 통해 단어 자질, 어절 자질, 구문 자질을 추출한다. 단어 자질은 내용어만 추출하고 어절 자질은 내용어와 정규화된 기능어(예를 들면 예게/한테 → 예게)를 자질로 추출하고 구문 자질은 의존어와 지배어 그리고 의존관계로 구성된다. 현재 자질 가중치는 정보검색에서 널리 사용되는 TF-IDF를 사용한다[11]. 본 논문에서 학생 답안을 하나의 문서로 간주하였다.

#### 3) 미채점 답안 군집화

[그림 2]는 미채점 답안의 군집화 알고리즘을 기술한다. 본 논문에서 제안된 군집화 알고리즘은 분류 확률과 분류 결과를 이용한다. 즉 분류 확률이 임계값 이상인 각 분류(class)를 하나의 군집으로 간주한다. 또한 분류의 정확도를 높이기 위해서 두 분류기의 결과가 같은 경우에만 정답으로 간주한다([그림 2]의 일곱 번째 단계를 참조하시오).

분류 확률을 계산하기 위해서는 로지스틱 회귀분석(Logistic Regression) 모델[12]을 사용하고 학생 답안의 분류를 위해서는 로지스틱 회귀분석 분류기와 k-NN(k-Nearest Neighbors) 분류기[12-13]를 사용한다. 특히 이는 학습 시간을 절약하기 위해서 로지스틱 회귀분석 분류기는 확률 계산과 분류의 목적으로 사용된다.

이 결과는 인터페이스를 통해서 인간 채점자에게 제공되고 인간 채점자는 의미적으로 유사한 학생 답안을 한꺼번에 확인할 수 있기 때문에 채점의 일관성을 유지하는데 큰 도움이 될 것이다. 인간 채점자가 군집 결과의 확인이 완료되면 그 결과는 학습 자료에 추가되어 새로운 학습 자료를 이용해서 채점 단계를 다시 수행한다.

```

1. 채점이 완료된 답안의 자질을 행렬(trainX)로 만든다.
2. // trainX를 이용해서 로그 회귀분석 분류기를 학습한다.
   clf1 = LogisticRegression(trainX, trainLabels)
3. // trainX를 이용해서 kNN 분류기를 학습한다.
   clf2 = KNeighborsClassifier(trainX, trainLabels)
4. 미채점 답안의 자질을 행렬(testX)로 만든다.
5. // 미채점 답안의 분류 확률을 계산한다.
   testProb = clf1.predict_proba(testX)
6. // 이미 학습된 두 분류기 clf1과 clf2를 이용해서 미채점
   답안을 채점한다.
   testy1 = clf1.predict(testX)
   testy2 = clf2.predict(testX)
7. // 미채점 답안을 군집화한다.
   nCorpus, nLabels = [], []
   for i in range(len(testy1)):
       if testProb[i][int(testy1[i])] > threshold
       and testy1[i] == testy2[i]:
           nCorpus.append(testCorpus[i])
           nLabels.append(testLabels[i])
8. return nCorpus, nLabels
    
```

[그림 2] 미채점 답안 군집 알고리즘

이와 같은 과정을 반복하면 학습 자료의 양은 증가하고 분류 확률의 신뢰성과 분류 정확률도 따라서 증가한다. 분류 확률의 신뢰성이 높아지므로 확률의 임계값을 점진적으로 감소하여 좀 더 빠른 속도로 미채점 답안을 채점할 수 있도록 구성되어 있다. 그러나 너무 빠른 속도로 감소되면 분류의 정확률이 떨어질 수 있으므로 인간 채점자가 임의로 조절할 수 있도록 하였다. 기본적으로 임계값은 0.99에서 출발하여 한 번 반복할 때마다 0.03씩 감소하도록 설정되어 있다.

#### 4. 실험 평가 및 분석

이 장에서는 제안된 서답형 문항 자동채점 시스템의 성능을 평가한다.

##### 4.1 실험에 사용한 서답형 문항 및 정답

평가에 사용된 서답형 문항은 2013년에 실시된 “국가수준 학업성취도 평가”의 사회 과목(중3)과 국어 과목(고2)에서 임의로 한 문항씩 선택한 것이다[14]. 각 문항은 1000개의 학생 답안과 채점 기준 답안으로 구성되어 있다.

##### 4.2 실험 결과

실험 결과는 1000개 답안을 사람이 직접 수동 채점한 결과(정답)와 자동채점 한 결과를 비교하여 분석하였다. 각 문항에 대한 자동채점 결과의 항목은 학생 답안의 수, 학생 답안의 유형, 채점 기준 답안의 수, 채점기준 점수의 수, 자동채점의 반복횟수, 채점자가 자동채점 시

시스템을 통해 주어진 학생 답안을 평가하는데 걸린 시간, 채점자가 수동채점을 통해 주어진 학생 답안을 평가하는데 걸린 시간, 자동채점과 수동채점의 차이(오류), 공백 답안을 포함한 자동채점의 정확률, 자동채점만 된 답안들의 정확률이다.

사회 과목(중3)과 국어 과목(고2) 서답형 항목의 자동채점 결과는 [표 1]과 같다.

[표 1] “국가수준 학업성취도 평가” 사회(중3), 국어(고2) 서답형 항목의 자동채점 결과

항목	자동채점 결과	
	사회(중3)	국어(고2)
학생 답안	1000	1000
학생 답안 유형	252	338
채점 기준 답안(정답)	1	1
채점기준 점수의 수	2	2
자동채점의 반복횟수	13	14
자동채점 총 걸린 시간(분)	11	35
수동채점 총 걸린 시간(분)	55	234
오류(수동채점≠자동채점 결과)	13	29
정확률(공백 답안 포함, 1000개)	98.7 %	97.2 %
정확률(자동 채점된 문항)	96.8 %	94.8 %

##### 4.2.1 사회 과목(중3) 서답형 문항

선택한 사회 과목(중3)의 서답형 문항은 ‘왕권신수설’의 의미를 작성하는 것이다. 해당 문항의 정답은 “왕권은 신이 부여한 것이다”이다.

[표 1]의 사회 과목의 채점 결과를 살펴보자. 먼저 서답형 문항 1000개의 답안을 채점하는데 걸린 시간을 비교하면 수동채점과 자동채점에 걸린 시간은 각각 55분과 11분으로 자동채점이 수동채점에 비해 80% 정도 시간을 절약했다. 수동채점의 결과가 완벽하다고 가정했을 때, 자동채점 결과에서 13개의 오류가 나타났다. 공백 답안을 포함한 자동채점의 정확률은 98.7%이며, 자동채점 된 407개의 항목에 관해서는 96.8%의 정확률을 보였다.

##### 4.2.2 국어 과목 서답형 문항

선택한 국어 과목(고2)의 서답형 문항은 ‘동물쇼 폐지’를 설득하는 글을 쓰기 위해 제시한 <개요>의 내용을 보고 ‘동물쇼 유지론에 대한 반론’으로 적절한 문장을 작성하는 것이다. 해당 문항의 모범답안은 “동물쇼에 동원된 동물은 야생 상태의 모습이 아니므로, 자연 생태 교육의 효과가 없다”이다.

[표 1]의 국어 과목 채점 결과를 살펴보자. 먼저 서답형 문항 1000개의 답안을 채점하는데 걸린 시간을 비교하면 수동채점과 자동채점에 걸린 시간은 각각 234분과 35분으로 자동채점이 수동채점에 비해 85% 정도 시간을 절약했다. 수동채점의 결과가 완벽하다고 가정했을 때, 자동채점의 결과 29개의 오류가 나타났다. 공백 답안을 포함한 자동채점의 정확률은 97.2%이며, 자동채점 된 539개의 항목에 관해서는 94.8%의 정확률을 보였다.

## 5. 결론 및 향후연구

본 논문에서는 학생들의 종합적인 사고능력을 판단하기 위한 서답형 문항의 채점에 드는 시간과 비용을 줄이기 위한 자동채점 시스템을 제안하였다. 먼저 언어 처리 기법을 이용하여 학생들의 답안을 분석 및 처리한 뒤, 이 결과가 일치하는 답안들을 묶어 고빈도 순으로 정렬하여 채점자에게 채점하게 하였다. 이렇게 채점된 결과에서 자질을 추출하여 자질의 가중치 학습을 한 뒤, 미채점 답안들에 로지스틱 회귀분석과 kNN 알고리즘을 이용하여 점수를 부여하는 방법을 사용하였다. 그 결과 1000개의 답안에 대한 채점시간이 수동채점 했을 때보다 최소 80%정도 감소했다. 수동채점이 완벽했다고 가정했을 때, 사회 과목과 국어 과목의 서답형 문항의 전체 정확률은 각각 98.7%와 97.2%로 나타났으며 자동채점으로 채점된 답안들의 채점 정확률은 각각 96.8%와 94.8%로 나타났다.

향후에는 본 논문에서 제안한 자동채점 시스템의 신뢰성을 높이기 위한 방법을 연구하여 적용할 계획이며, 사용자의 편의성을 위해 채점할 때 필요한 키워드를 제공하는 부분과 사용자가 직접 유의어 사전을 등록하고 편집할 수 있도록 하는 인터페이스를 추가하고 대규모의 답안에도 적용할 수 있도록 본 시스템의 시스템 구조를 서버-클라이언트 시스템에 적용할 예정이다.

## 참고문헌

- [1] 교육과학기술부, '2011년 국가수준 학업성취도 평가' 기본계획 발표, 보도자료, p. 2, 2011.
- [2] 이양락, 조지민, 신일용, 조운동, 이명애, 선택수, 박기범, 이광상, 김용명, 강대현, 김동영, 김현경, 김진구, 김영춘, 2014학년도 대학수학능력시험체제 개발을 위한 기초 연구, 한국교육과정평가원, p. 41, 2010.
- [3] 김종성, 국어과 서답형 평가 길라잡이, 충청남도교육청 중등교육정책과, p. 7, 2009.
- [4] 진경애, "영작문 자동 채점 시스템 개발 연구", 한국교육과정평가원, 영어어문교육, 제13권, 제1호, pp. 236-239, 2007.
- [5] Y. Attali and J. Burstein, "Automated essay scoring with e-rater", The Journal of Technology, Learning, and Assessment, vol. 4, no. 3, pp. 12-15, 2006.
- [6] M. D. Shermis and J. Burstein, Automated Essay Scoring: A Cross-Disciplinary Perspective. Lawrence Erlbaum Associates, Inc., Publishers. Mahwah, New Jersey, 2003.
- [7] O. Chapelle and B. Scholkopf, et al, Semi-Supervised Learning, Cambridge, England, MIT Press, pp. 1-8, 2006.
- [8] 국립국어원, 문화체육관광부, 21세기 세종계획, <http://korean.go.kr/sejong>, 2011.
- [9] 김성용, Tabular parsing 방법과 접속정보를 이용한 한국어 형태소 분석기, 한국과학기술원 석사 학위논문, pp. 21-37, 1987.

- [10] J. Hall, J. Nilsson and J. Nivre, MaltParser, <http://www.maltparser.org/>
- [11] G. Salton and M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, pp. 118-120, 1983.
- [12] G. Casella, S. Fienberg and I. Olkin, An Introduction to Statistical Learning with Applications in R, Springer, 2013.
- [13] P. Harrington, Machine Learning in Action, Manning Publications, pp. 18-37, 2012.
- [14] 한국교육과정평가원, 2013년도 국가 수준 학업 성취도 평가 기출답안 및 정답, <http://kice.re.kr/board.do?boardConfigNo=112&menuNo=372>, 2013.