학생 답안 정보를 활용한 반자동 정답 템플릿 구축 도구

장은서○, 강승식
국민대학교 컴퓨터공학부
{akdangz, sskang}@kookmin.ac.kr

Semi-automatic Answer Template Constructor using Student Responses

Eun-Seo Jang○, Seung-Shik Kang
Kookmin University, School of Computer Science

요 약

대학수학능력시험이나 학업성취도평가와 같은 중요한 시험은 다단계의 복잡한 과정을 거쳐 채점되며, 대규모의 인적·물적 자원이 투입된다. 컴퓨터 시스템을 활용하여 채점 과정에 소모되는 자원을 절감하고 정확성을 향상시키는 등 채점 효율의 향상을 목표로 하는 연구가 진행 중에 있다. 하지만 채점하려는 문항에 대한 정보 없이 시스템이 자동으로 정·오답을 판단하는 것은 불가능하며, 채점자가 문항에 대한 정보를 일정한 형식으로 가공하여 시스템에 제공하여야 한다. 본 논문은 채점하려는 학생 답안을 정보로 이용하여 시스템이 인지할 수 있는 정답 템플릿 형식으로 구축하는 방법을 소개한다. 또한 소개한 방법으로 작성된 정답 템플릿을 이용하여 자동 채점을 수행하였을 때의 채점 결과 및 통계 자료를 수록하였다.

주제어: 자동 채점, 정답 템플릿, 한글 처리, 자연어 처리

1. 서론

자연어 처리 기술을 이용하여 특정 문항의 학생 답안들을 자동으로 채점하는 방법은 관련 학계의 주요 연구 주제 중 하나이다. 특히 학업성취도평가 또는 대학수학능력시험과 같은 시험은 학생의 진로를 결정하는 중요한 시험이며, 학생 뿐 아니라 학부모 및 교사 등의 관심이 집중된다. 그렇기 때문에 학생 답안의 채점 또한 복잡한 다단계의 과정을 거쳐 신중하게 진행되며 이를 위해 대규모의 인적·물적 자원이 투입된다. 이러한 채점 과정의 일부를 자동화된 컴퓨터 시스템으로 대체하여 비용의 절감과 함께 채점의 정확도와 같은 효율을 향상시키는 방법이 연구 중에 있다.[1]

하지만 채점 시스템이 문제를 자동으로 인식하여 학생 답안의 정·오답 여부를 가리는 것은 현실적으로 불가능하며, 채점자가 문항을 채점하는 데 필요한 정보를 시스템이 인식할 수 있는 형식으로 가공하여 제공하여야 한다. 하지만 채점자는 채점 대상 과목의 전문가이지만 컴퓨터 시스템을 다루는 데에는 미숙할 수 있으며 시스템이 인식할 수 있는 형식의 정답 템플릿을 작성하는 데 불필요한 시간이 소모될 수 있다.

본 논문은 채점하려는 학생 답안을 먼저 분석하여 문항 채점에 필수적인 정보와 채점자가 부가적으로 작성한 채점 옵션, 모범 답안 정보, 항목 별 부여 점수 등의 정보들을 포함하는 완성된 채점 정보를 시스템이 인식 가능한 형식의 정답 템플릿 파일로 자동 구축해 주는 방법을 소개한다. 또한 소개한 방법을 이용하여 9개 문항에 대한 정답 템플릿을 생성하는 실험을 진행하였으며 이에 대한 채점 결과 및 통계 자료들을 수록하였다.

2. 관련 연구

2.1 자동 채점 시스템

자동 채점 시스템이란 특정 문항에 대한 학생 답안들을 각각 검사하여 점수를 부여하는 작업을 자동으로 수행하는 시스템이다. 그러나 채점 시스템이 문항을 분석 및 인식하여 학생 답안의 정·오답 여부를 가리는 것은 불가능하며, 채점자가 문항의 채점 기준들을 시스템에 제공해야 한다. 이러한 문항 정보는 자동 채점 시스템마다 상이하며 본 논문에서는 현재 개발 중인 KASS 2.0 자동채점 시스템(가제)에 적용할 수 있는 형태의 문항 정보 형식인 정답 템플릿을 반자동으로 구축할 수 있는 방법을 소개한다.

2.2 정답 템플릿

KASS 2.0의 정답 템플릿은 채점 옵션 정보, 모범 답안 정보, 고빈도 답안 수동 채점 정보, 개념기반 채점 정보의 4가지 기본 정보와 채점자의 선택에 따라서 추가적으로 단서어 정보를 저장하는 답안 정보 파일이다.[2]

채점 옵션 정보는 각 학생 답안의 전처리 옵션들을 저장한다. 선택 가능한 옵션으로는 특수 기호 제거, 띄어쓰기 오류 교정, 맞춤법 오류 교정, 불용어 제거가 있다. 선택한 전처리 옵션은 정답 템플릿을 만드는 과정뿐 아니라 실제 채점이 될 때에도 적용되며 각 문항의 특성에 따라 알맞은 옵션을 부여하여야 한다.

모범 답안 정보는 문항의 예시 정답이라고 할 수 있는 하나 이상의 문자열 집합이며, 모범 답안 점수는 문항에서 받을 수 있는 최고점을 의미한다. 고빈도 답안 정보는 특정 고빈도 답안 유형이 출현했을 때 부여할 점수 정보를 포함한다. 학생 답안들 중에서 정답으로 인정할 수 있는 핵심 개념을 공통적으로 포함하는 다양한 유형을 개념 정보로 만들어 처리할 수 있으며 개념기반 채점 정보는 이러한 개념 정보들을 저장한다. 단서어 목록은 정답 작성을 위해 필수적으로 포함되어야 하는 단어들의 목록이며 답안에 단서어 목록의 단어가 하나도 등장하지 않는 경우 해당 답안을 오답 처리한다.

3. 정답 템플릿 반자동 구축 도구

3.1 학생 답안 정규화

학생 답안을 분석하는 작업에 앞서 문항의 채점 요건에 맞게 학생 답안을 정규화하는 작업이 선행되어야 한다. 맞춤법, 띄어쓰기 또는 구두점 사용의 오류를 검사해야 하는 문항이 있는 반면, 그렇지 않은 문항도 존재한다. 채점자가 문항의 특성에 맞게 정규화 옵션을 설정하고 학생 답안을 불러오도록 하면 시스템은 불러온 학생 답안을 대상으로 정규화 작업을 수행한다. 표 1은 학생 답안을 정규화하는 예제를 보여준다.

표 1 학생 답안 정규화 예제

모범 답안	다만 그 성음을 바탕으로 이치를 다한 것 뿐이다.
학생 답안	Ⓐ: 다만 그 성음을 바탕으로 이치를 다한것뿐이다.
특수기호 제거	다만 그 성음을 바탕으로 이치를 다한것뿐이다
띄어쓰기 교정	다만 그 성음을 바탕으로 이치를 다한 것 뿐이다

정규화가 완료된 학생 답안들은 유형 별로 빈도를 조사한 다음 빈도 순으로 내림차순 정렬되어 채점자에게 보여진다. 학생 답안 유형 목록은 유형 별 빈도와 누적 빈도를 제공하며, 채점자는 이 목록을 이용하여 문항의 채점을 위한 부가 정보를 기입한다.

3.2 부가 정보 입력

채점자가 입력 가능한 부가 정보는 모범 답안 목록 및 점수, 고빈도 답안 부여 점수, 개념 생성 대상 및 개념 출현 시 부여 점수, 단서어 목록이 있다.

채점자는 문항 별로 제공되는 채점 기준을 이용하여 모범 답안 목록과 모범 답안 점수를 부여하고 고빈도 답안 유형 중에서 선택적으로 수작업 채점을 수행한다. 누적빈도 정보를 참고하여 어느 정도의 학생 답안이 채점될 수 있는지 확인할 수 있다. 학생 답안 유형 중에서 개념 정보를 생성할 필요가 있는 유형을 선택 및 생성하고 필요에 따라 단서어 목록을 작성한다.

표 2는 정답 템플릿 구축 도구가 학생 답안을 분석 및 제공하는 답안 유형 목록을 검토하여 채점에 필요한 정보를 작성한 예시를 보여준다. 표 2의 예제를 통해 ans_2~ans_5를 수작업 채점하였으며 ans_2, ans_4, ans_5를 이용하여 개념을 생성하는 것을 알 수 있다.

표 2 부가 정보 기입 예시

개념 생성	답안 유형	빈도	누적 빈도	점수
	ans_1	f_1	$\sum_{i=1}^{1} f_i$	
√	ans_2	f_2	$\sum_{i=1}^{2} f_i$	s_2
	ans_3	f_3	$\sum_{i=1}^{3} f_i$	s_3
√	ans_4	f_4	$\sum_{i=1}^{4} f_i$	s_4
√	ans_5	f_5	$\sum_{i=1}^{5} f_i$	s_5

3.3 정답 템플릿 구축

정답 템플릿은 json 형식으로 구조화하였고 정규화 옵션, 모범 답안 정보, 고빈도 답안 수작업 채점 정보, 개념 정보와 채점자의 선택에 따라 단서어 정보가 저장된다. 채점자가 부가 정보의 입력을 모두 마친 다음 정답 템플릿 구축 명령을 내리면 시스템은 학생 답안 분석 결과와 입력된 부가 정보를 결합하여 정답 템플릿을 작성하게 된다.

모범 답안과 고빈도 답안 채점은 문자열 완전 일치 방법을 이용하기 때문에 각 목록의 문자열들을 정답 템플릿에 저장하고, 각각 부여된 점수를 저장한다. 단서어 또한 문자열 정보만 필요하기 때문에 채점자가 기입한 문자열 집합을 그대로 파일에 저장한다. 개념은 생성 대상인 답안 유형의 문자열을 형태소 분석하여 각 어절을 어휘 토큰과 문법 토큰으로 분리하여 저장한다.

4. 실험 및 결과

4.1 실험 환경

자동 채점을 수행하기 위하여 KASS 2.0 을 이용하고 정답 템플릿은 자동 채점 시스템이 인식할 수 있는 형식으로 저장한다. 정답 템플릿 생성 및 자동 채점 실험을 수행하기 위해 9개 문항의 학생 답안들을 이용했다. Q_1~Q_5는 1개 어절로 정답을 서술할 수 있는 문항이고, Q_6~Q_9는 1개 문장으로 정답을 서술할 수 있는 문항이다. 정답 템플릿은 개발된 반자동 생성 도구에 의해 만들어진 그대로 별도의 수정 없이 사용하였다.

4.2 실험 결과 및 평가

동일 기준으로 생성했을 때의 문항 별 차이를 알아보기 위해 1개의 모범 답안을 입력하였고 빈도 10 이상인 답안 유형을 수작업 채점하였으며 빈도 3~9인 답안 유형을 개념으로 생성하였다. 표 3은 위의 기준을 이용하여 정답 템플릿을 생성한 결과이다.

표 3 정답 템플릿 생성 결과

문항	개수			
	전체	모범답안	고빈도답안	개념
Q_1	3013	1	6	6
Q_2	3018	1	10	11
Q_3	3013	1	10	8
Q_4	3023	1	8	4
Q_5	3035	1	4	4
Q_6	3013	1	20	35
Q_7	3012	1	5	7
Q_8	3010	1	17	23
Q_9	3010	1	27	25

표 4는 생성한 정답 템플릿을 이용하여 실제 자동 채점을 수행한 결과이다. 실험 결과, 비교적 간단한 문항인 Q_1~Q_5 문항의 경우 학생 답안의 대부분이 모범 답안과 고빈도 답안 정보를 이용한 완전일치 채점에서 처리됨을 확인하였다. 1개 문장으로 이루어진 Q_6~Q_9 문항의 경우에는 Q_1~Q_5 문항보다 개념 기반의 채점 빈도가 늘어났으며

미판단 답안의 비율 또한 약간 증가하였다. 미판단 답안의 비율은 Q_1~Q_5 문항의 경우 평균 약 1.3%이고 Q_6~Q_9 문항은 평균 약 4.1%이다.

표 4 자동 채점 수행 결과

문항	개수				
	전체	모범답안 완전일치	고빈도답안 완전일치	개념 기반	미판단 답안
Q_1	3013	2806	109	25	73
Q_2	3018	2603	296	66	53
Q_3	3013	2611	318	33	51
Q_4	3023	2249	722	21	31
Q_5	3035	1774	1229	23	9
Q_6	3013	67	2619	199	128
Q_7	3012	2004	858	30	120
Q_8	3010	807	1982	91	130
Q_9	3010	816	1945	130	119

5. 결론 및 향후과제

본 논문은 자동 채점 시스템의 원활한 동작을 위해서 학생 답안을 분석 및 정보를 가공하고 채점자가 작성한 부가 정보를 이용하여 문항 채점 정보 파일인 정답 템플릿을 생성하는 방법을 소개하였다. 또한 소개한 방법으로 프로그램을 작성하여 9개 문항에 적용하는 실험을 진행하였다. 생성된 정답 템플릿을 이용하여 채점한 결과 1어절~1문장으로 구성된 학생 답안의 약 96.5%를 채점하였다.

하지만 여전히 미판단 답안이 존재하고 이로 인해 채점 작업을 한 번 이상 중복 수행해야 하는 문제가 있다. 향후 정답 템플릿 생성 과정을 보완하여 한 번의 채점 사이클로 모든 학생 답안을 처리할 수 방법에 대한 연구를 진행할 예정이다. 현재 시행중인 다양한 시험에 자동 채점 시스템을 적용하여 보다 효율적인 채점 작업이 가능할 것으로 기대한다.

참고문헌

[1] 성태제, 양길석, 강태훈, 정은영, “학업성취도평가 서답형 문항 컴퓨터 채점화 방안 탐색,” 한국교육과정평가원 연구보고서 RRE 2010-1, 2010.

[2] 박일남, 노은희, 심재호, 김명화, 강승식, “한국어 서답형 자동채점을 위한 정답 템플릿 기술 방법,” 제 24회 한글 및 한국어 정보처리 학술대회, pp.138-141, 2012.

[3] 박은아, 이문복, “컴퓨터기반 문제해결 능력 평가 시범 적용과 향후 과제,” 한국교육과정평가원 연구보고서 ORM 2012-92, 2012.

[4] 강승식, 한국어 서답형 문항(단어-구 수준) 자동채점 프로그램 개발, 한국교육과정평가원 연구보고서, 2012.

[5] 강승식, 한국어 형태소 분석과 정보 검색, 홍릉과학출판사, 2002.

[6] 노은희, 심재호, 김명화, 김재훈, "대규모 평가를 위한 한국어 서답형 자동채점 프로그램 개발의 전망," 한국교육과정평가원 연구자료 ORM 2012-92, 2012.

[7] S. Dikli, "An overview of automated scoring of essays," The journal of technology, learning, and assessment, pp.1-35, 2006.