

질의문 유형 분석을 통한 서답형 자동 채점 시스템

Automatic Grading System for Subjective Questions Through Analyzing Question Type

강원석

안동대학교 정보과학교육과

Won-Seog Kang(wskang@andong.ac.kr)

요약

서답형 자동평가는 자연어 처리의 어려움이 있어 그 시스템의 개발이 쉽지 않다. 본 연구는 자연어처리 기술을 접목한 서답형 자동채점 시스템을 설계, 구현한다. 이 시스템은 일반화된 채점 시스템이 지니는 성능 저하의 문제를 해결하기 위해 문제유형을 정의하고 각 유형에 맞는 처리를 하여 성능을 향상하였다. 이 시스템의 성능을 실험하기 위하여 교사 채점과 기존의 용어중심의 채점 시스템, 교사채점과 제안한 채점 시스템의 상관계수를 분석하였다. 실험결과, 기존의 용어 중심의 채점 시스템보다 향상된 결과를 얻었다. 앞으로 문제 유형을 확대 정의하고 각 유형에 맞는 자연어 처리 기법을 개선할 필요가 있다.

■ 중심어 : | 서답형 평가 | 자연어 처리 |

Abstract

It is not easy to develop the system as the subjective-type evaluation has the difficulty in natural language processing. This thesis designs and implements the automatic evaluation system with natural language processing technique. To solve the degradation of general evaluation system, we define the question type and improve the performance of evaluation through the adaptive process for each question type. To evaluate the system, we analyze the correlation between human evaluation and term-based evaluation, and between human evaluation and this system evaluation. We got the better result than term-based evaluation. It needs to expand the question type and improve the adaptive processing technique for each type.

■ keyword : | Subjective-type Evaluation | Natural Language Processing |

1. 서론

평가는 교육의 중요한 요소로 다양한 방법이 제공되고 있다. 선택형 평가는 신뢰성은 높으나 학생들의 고등정신능력을 평가하기에는 부족하다. 반면 서답형 평가는 선택형 평가의 단점을 보완하나 신뢰성의 문제가 있다. 이와 같은 문제를 해결하기 위하여 서답형에 대한 자동

채점 시스템의 연구가 이루어져 왔다.

[1-3]은 서답형 평가의 자동채점 시스템의 필요성을 인식하여 영작문 등의 영어에 대한 자동채점 시스템의 도입과 적용에 대한 연구를 하였다. 이 연구들에서 영어에 대한 서답형 자동채점의 연구는 많이 진행되었으나 우리말에 대한 서답형 평가는 아직 연구가 미흡하다고 말하고 있다.

[4-6]에서는 영어 논술답안을 자동으로 평가하기 위하여 다양한 자연어처리 기술을 적용하였다. 이 기술은 영어를 위한 것이라 한국어 평가에 적용할 수가 없다. [7-9]는 우리말에 대한 서답형 평가의 연구를 하였다. [7]은 자연어에 적용하기에는 한계가 있는 형식문법을 이용하여 구문구조를 도출하였고 이를 한문 영역에 적용하였다. [8][9]는 워드넷을 이용하여 의미적으로 유사한 용어를 함께 추출하는 방법을 적용하였다. 이와 같은 연구들은 질의에 대해 일반화된 채점방법을 적용한 것이다. 따라서 각 질의 유형에 따른 독특한 처리기법을 적용하지 못하여 전체 성능이 저하되는 문제가 있다. 본 연구는 이를 해소하기 위하여 질의 유형을 분석, 정의하고 각 질의 유형에 따른 적응형 처리를 하는 서답형 자동채점 시스템을 제안한다.

본 연구의 2장에서는 본 시스템에 대한 관련 연구를 서술하고 3장은 질의유형 분석을 통한 서답형 자동채점 시스템을 서술한다. 4장에서는 제안한 시스템의 성능을 분석하기 위하여 교사 채점결과와 본 시스템의 채점 결과의 상관관계를 분석하고 검토한 후 5장에서 결론을 지었다.

II. 관련 연구

1. 서답형 평가

서답형 평가는 단편적이고 지엽적인 지식보다는 종합적이고 고차적인 사고능력을 요구하는 평가에 적합하다. 그렇지만 채점자의 주관성으로 인한 신뢰도의 저하 때문에 객관성이 있는 자동 채점의 필요성이 있다. 이 절에서는 서답형 자동 채점을 위해 서답형 문항의 특징에 대해 살펴본다.

서답형 평가는 그 유형에 따라 다양하게 분류된다. [10]에 의하면 서답형 평가문항은 단답형, 괄호형, 완성형, 서술형(논술형)으로 구분한다. 단답형은 단어, 구, 절 혹은 수나 기호 등으로 답하는 유형이다. 간단한 용어의 의미나 계산 문제에 사용된다.

괄호형은 질문의 문장에 여백을 두고 이를 채우는 형태의 문항을 말한다. 완성형 문항은 질문의 마지막에 응

답을 하게 하는 유형으로 괄호형의 특수한 형태로 정의한다.

서술형은 문장형태로 답을 하는 유형이지만 객관적인 답이 있는 형태이고 논술형은 논리성, 창의성 등을 평가하기 위해 논술식으로 답을 하도록 하는 형태이다. 본 연구에서는 자연어로 된 문장을 서술하는 방식인 서술형과 논술형을 따로 구분하지 않고 서술형으로 용어를 통일하여 사용한다.

2. 자연어 처리기법

자연어 처리(Natural Language Processing)란 사람이 사용하는 자연언어를 분석하여 응용 영역에 따라 사용하는 것이다. 자연언어 분석은 일반적으로 형태소 분석, 구문 분석, 의미 분석, 담화 분석 등의 단계로 구성된다 [11]. 자연어 처리 기술은 언어를 사용하는 다양한 영역에 사용되나 영역에 따라 적용 기술이 다르다. 정보검색 분야의 경우 검색에 주요한 용어를 추출하기 위해 형태소 분석 기술을 사용하고 있고 기계번역의 경우는 구문 분석, 의미 분석의 단계까지 기술을 응용한다[12]. 서답형 평가 분야에는 적용 사례가 아직 적지만 형태소 분석의 기술을 적용할 수 있고 앞으로 구문 분석과 의미 분석의 기술이 응용되리라 생각된다. 본 연구에서는 서답형 평가에 적합한 자연어 처리기술을 찾기 보다는 서답형 유형의 분석과 그 유형에 따른 적응형 처리를 하는 것에 초점을 맞춘다. 따라서 이 연구에 사용된 자연어 처리 기술은 기본적인 형태소 해석 기술을 사용한다.

3. 선행 연구

서답형 평가 시스템은 자연어 처리의 어려움으로 그 연구가 많지 않았다. 최근 자연어 처리기술을 이용한 다양한 연구가 진행되고 있다.

[1][3]은 영작문의 서답형 평가를 위한 자동채점 프로그램의 도입 방안과 활용가능성을 모색한 연구이다. 이 연구는 영어와 함께 한국어 서답형 자동채점의 필요성을 제시하였다. [2]도 역시 영어 서답형 프로그램 적용과 과목별 서답형 채점 프로그램의 적용 가능성을 연구한 것이다. 이 연구 역시 한국어 서답형 자동채점의 필요성을 제시하고 있다.

[5]는 영어의 다양한 서답형 평가 시스템에 대한 비교, 분석을 하였고 [4][6]은 영어 논술의 채점 방법을 제안하였다. 그렇지만 이 기술들은 영어에 대한 것이라 한국어 서답형 평가에 적용할 수가 없다.

[13][14]는 시스템을 통해 서답형 평가를 공정하게 할 수 있도록 하는 보조 시스템을 제안하였다. [10]은 서답형 문항 평가에 대한 설계 모델과 평가 방법 등을 제시하고 있다. 이 연구들은 서답형 문항을 자동으로 채점하는 시스템에 대한 필요성과 방향을 제시하였다.

[9]는 자연어 처리기법을 이용하여 선택형 문항을 자동으로 생성하는 연구를 하였고 [15]는 형식언어인 프로그래밍 언어에 대한 서답형 자동채점 시스템에 대한 연구를 하였다. 이 연구들은 서답형 자동채점에 대한 관련 기법을 제안하였다.

[16]은 주관식 채점 모형의 필요성과 문제점을 제안하고 서답형 유형 가운데 단답형과 완성형에 대한 유형을 부분점수와 불리안 연산을 통해 채점을 한 연구이다. 이 연구는 서답형 평가의 주요 유형인 서술형 유형에 대한 처리가 미흡하다.

[7]은 자연어에 적용하기에는 한계가 있는 형식문법의 파싱기법을 적용하여 구문트리를 얻었고 이를 이용하여 한문 분야에 적용한 것이다. 이 연구는 형식문법의 적용과 용어 비교 또는 구조 비교에서 정확한 매치만 인정하는 등의 한계가 있다.

[17]은 유의어 사전을 이용하여 서답형 평가를 한 것으로 제한적인 사전을 이용한 한계가 있다.

[8]은 한글 워드넷을 이용하여 유의어를 추출하고 의미커널을 구축한 후 유사도 비교를 통해 채점하는 시스템이고 [18]은 형태소 분석에만 의존하던 방식에서 구문 해석을 위해 트리형태를 구성하여 가중치를 주고 이를 비교하여 채점하는 시스템으로 좋은 결과를 얻었다. 그러나 이 연구들은 자연어처리기법을 적용하여 좋은 결과를 얻었으나 서답형의 다양한 질의 유형을 인지하지 못하여 성능 저하의 문제가 발생할 수 있다. 이에 본 연구는 이 문제를 개선하고자 서답형 유형의 질의 유형을 분석하고 각 질의 유형에 따른 적용형 처리를 하는 서답형 자동채점 시스템을 제안한다.

III. 질의유형 분석을 통한 서답형 자동채점 시스템

1. 질의 유형

본 연구에서는 서답형의 다양한 예를 검토하여 [표 1]과 같은 질의 유형을 분류, 정의하였다.

표 1. 질의 유형 종류

질의 번호	질의명	예
1	기본형	플립플롭의 특징을 쓰시오.
2	공통과제형	집중화 장치의 개념과 특징을 쓰시오.
3	순서제시형	명령어 실행에서 간접사이클 과정을 설명하라.
4	설명제시형	다음 용어를 설명하시오.
5	장단점제시형	가상기억장치의 장단점을 기술하시오.
6	부분제시형	B-ISDN의 특징을 3가지 이상 쓰시오.

- (1) 기본형 : 하나의 논제를 제시하여 그 답을 요구하는 일반의 형식이다. 예를 들면 ‘스택의 특징을 쓰시오.’와 같은 질의 유형이다.
- (2) 공통과제형 : 하나의 논제에 두 가지 답을 작성하게 하거나 두 논제의 답을 작성하는 형식으로 AND 형식의 질의가 이에 해당한다. 예를 들면 ‘스택과 큐의 특징을 쓰시오.’와 같은 유형이다.
- (3) 순서제시형 : 순서나 절차에 대한 질의로 순서가 중요한 답안인 유형이다. 예를 들면 ‘컴파일러의 처리과정 5단계를 기술하라.’와 같은 유형이다.
- (4) 설명제시형 : 어떤 용어나 진술에 대한 정확한 뜻이나 분류, 그 용어의 설명을 요구하는 유형이다. 예를 들면 ‘컴파일러의 종류를 나열하고 설명하시오.’와 같은 유형이다.
- (5) 장단점제시형 : 논제에 대해서 장점과 단점을 요구하는 유형이다. 예를 들면 ‘해싱 기법의 장단점을 설명하시오.’와 같은 유형이다.
- (6) 부분제시형 : 논제의 답이 여러 가지가 있을 때 그 중 몇 가지만 요구하는 유형이다. 예를 들면 ‘프로그래밍 언어의 종류를 3가지 이상 쓰시오.’와 같은 유형이다.

위의 유형들 가운데 두 개 이상의 유형으로 분석될 수도 있다. 그러한 경우 제안한 질의 유형 결정 알고리즘에 따라 선택한다. 최종적으로 어디에도 해당하지 않는 경

우는 기본형으로 정의한다.

2. 서답형 자동채점 시스템 구조

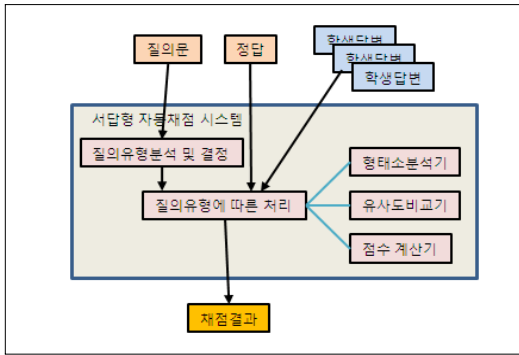


그림 1. 서답형 자동채점 시스템 구조도

본 연구의 서답형 자동채점 시스템은 [그림 1]과 같이 구성되었다. 시스템은 먼저 질의문을 받아 질의 유형을 분석하여 질의 유형을 결정한다. 다음으로 정답과 학생답변을 받아 형태소 분석기, 유사도 비교기, 점수 계산기의 도움을 받아 질의유형에 따른 자동 채점을 실시하여 그 결과를 내어놓는다.

형태소 분석기는 [19]의 시스템을 이용하였다. 그 결과에서 용어를 추출하고 용어의 빈도수를 계산하여 정답과 학생답들의 벡터 표현을 구한다. 각 벡터 표현은 다음과 같이 정의된다.

$$A = (a_1, a_2, a_3, \dots, a_n), \quad (1)$$

$$a_j = \frac{freq_j}{\sum_{k=1}^n freq_k}$$

$$S_i = (s_{i1}, s_{i2}, s_{i3}, \dots, s_{im}), \quad (2)$$

$$s_{ij} = \frac{freq_{ij}}{\sum_{k=1}^m freq_{ik}}$$

A는 정답의 표현이고 S_i는 i번째 학생답의 표현을 나타낸다. a_j와 s_{ij}는 용어의 가중치 값으로 빈도수를 이용하여 정의한다. n는 정답에서 나타난 용어의 개수이고 m은 각 학생답에서 나타난 용어의 개수이다.

유사도 비교기는 정답과 학생답들의 표현의 벡터값을

이용하여 그 유사한 정도를 검사하는 것이다. 여러 가지 유사도 비교함수가 있으나 본 연구에서는 널리 사용되는 코사인 함수를 이용하였다. 그 정의는 다음과 같다.

$$sim(S_i, A) = \frac{S_i \cdot A}{|S_i| \times |A|} = \frac{\sum_{j=1}^p s_{ij} \times a_j}{\sqrt{\sum_{j=1}^p s_{ij}^2} \times \sqrt{\sum_{j=1}^p a_j^2}} \quad (3)$$

이 함수를 적용할 때 정답의 용어와 학생답의 용어가 서로 다르면 학생답의 용어집합과 정답의 용어집합의 합집합을 구하여 전체 용어집합으로 정의한 후 적용한다. p는 전체 용어집합의 원소 개수가 된다.

점수 계산기는 질의 유형에 따라 별도의 점수를 계산하는 것으로 다음 절에 설명하는 질의유형 결정과 처리 알고리즘에서 정의된다.

3. 질의 유형 결정과 처리 알고리즘

질의문이 입력되면 질의 유형 결정 알고리즘에 따라 질의 유형을 결정한다. 질의 유형 결정 알고리즘은 다음과 같다.

- ① 질의문 형태소 분석
- ② 질의문의 용어와 각 유형의 휴리스틱 규칙 매치 검색
- ③ 중복 매치된 경우 threshold 값 이상의 차이가 나는 유형 선택
- ④ 그렇지 않은 경우 휴리스틱 선택 규칙 적용

첫 단계로 질의문을 형태소 분석을 한다. 그 결과 형태소 리스트가 출력된다. 둘째, 이를 이용하여 [표 2]의 질의유형 결정 휴리스틱 규칙을 이용하여 각 유형과의 매치 여부를 검사한다. 셋째, 중복 매치된 경우 threshold 값 이상의 차이가 나는 유형을 선택한다. 그렇지 않으면 휴리스틱 선택 규칙을 적용한다.

표 2. 질의 유형 결정 휴리스틱 규칙

질의 번호	질의유형	휴리스틱 규칙
1	기본형	default
2	공통과제형	과, 이랑, 하고 등의 병렬연결접속사 여부
3	순서제시형	과정, 단계, 절차, 순서, 순, 차례, 진행, 공정, 경로 등의 관련단어 여부
4	설명제시형	의미, 용어, 단어, 뜻, 개념, 종류, 분류 등의 관련단어 + 나열, 열거, 설명, 밝히다의 여부
5	장단점제시형	장단점, 장점, 단점, 좋은 점, 나쁜 점, 이점, 문제점, 이득, 득실 등의 관련 단어 여부
6	부분제시형	N가지, N개, N종류 (+ 이상)의 여부

휴리스틱 선택 규칙은 질의문이 제시한 질의 유형 가운데 해당되는 것이 두 가지 이상이고 한계값 이상으로 가려지지 않는 경우 우선순위가 높은 것을 선택하는 방법이다. 그 우선순위는 순서제시형, 설명제시형, 장단점제시형, 부분제시형, 공통과제형, 기본형이다. 이 우선순위는 순서에 관련된 용어는 다른 유형에 나타나지 않는 특징을 가진 순서제시형처럼 다른 것보다 변별성이 더 뚜렷하고 차별성이 높은 것이 높게 책정되었다.

질의 유형이 결정되면 각 질의 유형에 따른 처리를 한다. 각 경우의 처리 방법은 다음과 같다.

(1) 기본형

다른 질의 유형으로 판단되지 않는 경우 묵시적으로 처리하는 방법이다. 이 유형의 처리는 기존의 다른 연구처럼 정답과 학생답을 형태소 분석하고 이를 벡터로 표현한 후 정의한 유사도 함수를 이용하여 점수화한다.

(2) 공통과제형

병렬 연결접속사로 연결된 질의문의 경우에 해당한다. 이의 처리 단계는 다음과 같다.

- ① 질의문에서 대등 단어 검색
- ② 정답과 학생답의 영역 검색
- ③ 두 영역별 벡터표현한 후 비교하여 점수화
- ④ 두 영역의 점수의 합산값으로 최종점수화

첫 단계로 질의문에서 병렬연결접속사로 연결된 대등 용어를 검색한다. 예를 들면 ‘목적과 특징’의 경우 목적과 특징이 대등용어가 될 것이다. 둘째, 대등 용어를 이용하여 정답과 학생답에서 영역을 구분하여 두 영역으로 나뉜다. 만약 학생답에서 대등 용어를 이용하여 영역

을 구분하지 못하는 경우 다음의 클러스터링 기법을 사용하여 두 영역으로 구분한다. 클러스터간의 비교는 유클리디안 거리함수를 사용하였다.

$$Dist(A, B) = \sum_{k=1}^n (a_k - b_k)^2 \quad (4)$$

학생답을 두 영역으로 나누어 하나는 A, 다른 것은 B로 정의하자. 답변의 특성상 두 영역의 자료가 경계선을 통해 구분되므로 경계를 학생답변 형태소 해석의 결과인 용어 리스트의 첫 용어부터 시작하여 마지막 용어까지로 변화시켜가면서 A와 B의 거리를 구하여 가장 거리가 먼 경우를 경계로 정의하여 두 영역으로 구분한다 이것은 관련있는 용어들이 가까이 출현하는 locality 원칙을 이용한 것이다. 셋째, 두 영역에 대한 용어들을 벡터로 표현한 후 유사도함수를 이용하여 정답 영역과 비교하여 점수화한다. 넷째, 두 영역의 점수를 합산하여 최종 점수를 구한다. 두 영역의 점수비는 동등하게 주었다.

(3) 순서제시형

순서나 절차에 대한 질의문인 경우이다. 이 경우 다음의 처리 과정을 따른다.

- ① 정답과 학생답의 영역검색
- ② 정답의 N개 영역과 학생답의 M개 영역을 N*M개의 벡터표현의 비교후 정답 영역과 매치되는 학생답 영역 선택
- ③ 정답의 영역과 매치된 학생의 영역의 순서 정보 확인 후 점수화

첫 단계에서 숫자나 기호, 서수 등을 검사하여 정답과 학생답의 영역을 구분한다. 정답의 경우 N 개의 영역으로 구분된다고 하고 학생답의 경우 M 개의 영역으로 구분된다고 하자. 둘째, 정답의 각 영역에 대해 학생답 M개영역을 각각 벡터 표현후 이를 비교하여 가장 유사도가 높은 영역을 선택한다. 만약 선택된 영역의 유사도가 threshold 이상 되지 않으면 그 정답 영역에 해당하는 학생의 답이 없다는 것으로 정의한다. 셋째, 정답 영역 N 개에 대해 매치된 학생답 영역이 N개이면 옳게 답한 것이다. 정답 영역 N 개 가운데 일부 R 개만 매치되었다면 점수는 R/N으로 계산된다. 이 문항은 순서가 매우 중요

한 것이므로 매치된 것의 순서가 정답의 순서와 일치하면 순서정보점수를 부여한다. 만약 학생답이 일부만 정답순서의 부분순서를 유지한다면 그 부분순서를 지니는 P 개의 영역에 대해 P/N의 순서정보점수를 부여한다. 순서정보와 영역일치정보의 점수비는 동등하게 부여하였다.

(4) 설명 제시형

어떤 용어의 개념이나 종류 등을 열거하고 설명하는 유형이다. 이 경우 다음의 절차를 따른다.

- ① 정답과 학생답의 영역구분
- ② 정답의 N개 영역과 학생답의 M개 영역을 N*M개의 벡터표현의 비교후 정답 영역과 매치되는 학생답 영역 선택
- ③ 각 영역에 대한 점수 합산

첫 단계에서 정답과 학생답의 내용 가운데 영역을 구분할 수 있는 용어나 기호, 숫자 등을 이용하여 영역을 구분한다. 만약 구분되지 않는 경우 클러스터링 기법을 사용하여 구분한다. 이때 학생답의 크기가 평균 크기보다 클 경우 영역을 구분하였다. 둘째, 정답의 각 영역에 대해 학생답 M개영역을 각각 벡터 표현후 이를 비교하여 가장 유사도가 높은 영역을 선택한다. 셋째, 정답 영역에 매치된 학생답 영역의 유사도 비교 점수를 합산하여 최종 점수를 계산한다.

(5) 장단점 제시형

논제에 대해 장점과 단점을 묻는 유형이다. 이 경우 다음의 처리절차를 따른다.

- ① 정답과 학생답의 영역 검색
- ② 정답의 두 영역에 대해 학생답의 영역과 비교하여 매치되는 정도가 높은 영역을 선택
- ③ 두 영역의 점수의 합산

첫 단계에서 정답과 학생답의 내용 가운데 영역을 구분할 수 있는 용어나 관련단어 등을 이용하여 영역을 구분한다. 만약 구분되지 않는 경우 클러스터링 기법을 사용하여 구분한다. 둘째, 정답의 두 영역에 대해 학생답의 영역을 각각 비교하여 유사도가 높은 영역을 선택한다. 셋째, 정답 영역에 매치된 학생답 영역의 유사도 비교 점수를 합산하여 최종 점수를 계산한다.

(6) 부분 제시형

논제의 답이 여러 가지가 있을 때 그 중에 몇 가지만 요구하는 유형이다. 그 처리 절차는 다음과 같다.

- ① 정답과 학생답의 영역구분
- ② 정답의 N개 영역과 학생답의 M개 영역을 N*M개의 벡터표현의 비교후 정답 영역과 매치되는 학생답 영역 선택
- ③ 각 영역에 대한 점수 합산

첫 단계에서 정답과 학생답의 내용 가운데 영역을 구분할 수 있는 용어나 구분 기호, 숫자 등을 이용하여 영역을 구분한다. 만약 구분되지 않는 경우 클러스터링 기법을 사용하여 구분한다. 이때 학생답의 크기가 평균 크기보다 클 경우 영역을 구분하였고 본 연구에서는 M을 3 이하로 한정하여 실험하였다. 둘째, 정답의 각 영역에 대해 학생답 M개영역을 각각 벡터 표현후 이를 비교하여 가장 유사도가 높은 영역을 선택한다. 셋째, 정답 영역에 매치된 학생답 영역의 유사도 비교 점수를 합산하여 최종 점수를 계산한다.

IV. 실험 및 분석

제안한 서답형 자동채점 프로그램의 실험에 앞서 채점에 대한 기준을 제시하기 위하여 6명 교사들의 채점결과를 얻고 기본형을 제외한 5가지 유형에 대한 질의문의 학생답을 채점하였다. 이때 각 채점의 최고값과 최저값을 제외한 점수들의 평균값으로 기준을 삼았다.

본 연구에서는 세가지 유형의 시스템을 구축하여 실험한다. 그 시스템의 종류와 특징은 [표 3]과 같다. 본 연구에서 제안하는 시스템은 시스템 3에 해당한다.

표 3. 실험 시스템의 종류와 특징

시스템종류	특징
시스템 1	질의 유형을 구분하지 않고 기본형으로 처리하는 시스템
시스템 2	질의 유형을 구분하지만 클러스터링 기법으로 영역을 구분하여 처리하는 시스템
시스템 3	질의 유형을 구분하고 본 연구에서 제안한 각 유형별 처리 절차를 따르는 시스템으로 정답의 영역에 대한 학생답 영역의 최적매치 선택후 점수 계산하는 방법 적용함.

시스템 1은 [7]의 기법을 적용한 시스템으로 질의유형을 분석하지 않고 형태소 해석후 결과의 용어의 가중치를 이용하여 유사도 비교를 하여 점수화한 시스템이다. 시스템 2는 본 연구에서 제안한 기법에서 영역 구분에 대한 방법을 정보검색 분야의 클러스터링 기법을 적용한 시스템이다. 이 시스템은 질의 유형을 구분하지만 각 유형에 따라 영역 구분을 할 때 각 유형별 구분 기호나 단어를 이용하지 않고 클러스터링 기법으로 영역을 구분하여 처리하는 시스템이다. 시스템 3은 본 연구에서 제안한 시스템으로 질의 유형 분석에 따른 처리를 하는 시스템을 말한다.

세 시스템의 실험 결과를 분석하기 위해 평가 분야에서 많이 사용되는 피어선 적률 상관계수를 사용하였다. 상관계수는 채점자간의 신뢰도를 얻기 위한 것으로 두 채점자의 채점이 얼마나 일치되는지를 보여주는 것이다. 본 실험에서는 기준이 되는 교사 채점평균과 세 시스템간의 상관도를 구하여 어느 시스템이 교사채점과 더 상관도가 높은지를 알아보았다. 상관계수 식은 다음과 같다.

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad (5)$$

이 식의 \bar{x} 와 \bar{y} 는 채점의 평균을 나타내고 σ_x 와 σ_y 는 표준편차를 나타낸다.

각 유형별 검사 문항은 [표 4]에 기술하였고 각 문항별 교사 채점과 각 시스템에 대한 검사 결과는 [표 5]와 같다.

표 4. 유형별 검사 문항

문항번호	유형	검사문항
1	공통과제형	컴퓨터 통신의 목적과 특징에 대해 기술하십시오.
2	순서제시형	인터럽트의 동작 원리를 설명하십시오.
3	설명제시형	전파지연, 전송시간, 처리지연의 용어를 설명하십시오.
4	장단점제시형	프로토콜 데이터 단위(PDUP)를 작게 단편화할 경우 장단점을 쓰시오.
5	부분제시형	B-ISDN의 특징을 3가지 이상 기술하십시오.

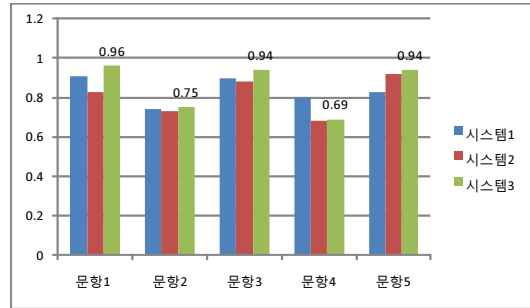


그림 2. 문항별 시스템 실험 결과 비교

표 5. 문항별 시스템 실험 결과

문항종류	시스템종류	학생1	2	3	4	5	6	7	8	9	평균	상관계수	
문항1	교사채점	52.50	60.25	23.00	42.50	90.00	52.50	83.75	61.00	88.75	61.58	0.91	
	시스템1	34.25	39.02	29.06	42.09	85.86	49.82	72.85	55.33	69.01	53.03		
	시스템2	30.92	32.57	34.26	10.74	76.10	40.34	73.66	38.33	61.53	44.27		0.83
	시스템3	58.54	59.81	38.38	50.76	87.56	58.58	99.07	64.89	89.15	67.41		0.96
문항2	교사채점	81.00	61.25	85.00	71.25	94.25	75.00	92.50	70.00	69.75	77.78	0.74	
	시스템1	64.78	63.69	80.05	64.68	83.31	80.12	79.31	60.50	71.90	72.04		
	시스템2	62.52	47.14	100	37.51	100	100	100	18.89	68.40	70.50		0.73
	시스템3	48.02	36.20	100	21.60	100	93.22	90.90	14.51	53.30	61.97		0.75
문항3	교사채점	20.00	48.75	97.50	32.50	100	33.75	88.75	85.00	97.50	67.08	0.90	
	시스템1	35.08	54.73	82.53	65.32	84.67	51.65	86.11	92.78	78.17	70.12		
	시스템2	26.24	28.87	80.78	51.61	85.35	45.68	71.91	87.74	70.25	60.94		0.88
	시스템3	34.11	37.53	100	47.94	100	59.38	93.48	100	91.32	73.75		0.94
문항4	교사채점	51.25	57.50	59.50	57.50	89.50	0.00	75.75	75.00	69.50	59.50	0.80	
	시스템1	45.89	26.47	71.76	54.56	89.30	6.16	77.81	45.89	89.12	56.33		
	시스템2	38.11	31.69	68.00	26.98	85.58	7.75	37.95	38.11	86.32	46.72		0.68
	시스템3	45.74	70.17	81.60	32.38	100	9.30	46.00	45.74	100	58.99		0.69
문항5	교사채점	60.00	46.75	86.25	15.50	100	85.00	90.00	41.25	87.50	68.03	0.83	
	시스템1	55.43	51.64	70.80	28.40	84.95	43.88	66.20	38.85	59.43	55.51		
	시스템2	55.43	30.05	79.59	20.82	82.96	58.53	57.97	38.85	66.96	54.57		0.92
	시스템3	55.43	36.06	95.51	24.99	99.55	70.23	69.57	38.85	80.35	63.39		0.94

[그림 2]의 그래프는 교사 채점과 3개의 시스템 결과간의 상관계수를 잘 보여주고 있다. 4개의 문항에서 비교적 시스템 3이 나은 성능을 보여 주고 있다. 그러나 2번과 4번의 문제유형은 시스템 1이 높게 또는 유사하게 나왔다. 이는 영역 구분과 점수 반영 방법 등이 효과적으로 이루어지지 못한 것으로 보인다. 특히 문항 2번의 경우 순서 정보가 있어 이를 효과적으로 반영할 방법의 개선이 필요하다고 분석된다. 문항 4의 경우 영역을 구분한 것이 오히려 좋지 않은 결과를 가져온 것으로 판단된다. 실제 학생 답안이 장점과 단점이 혼재하여 표기한 경우도 많았다. 이 문제를 해결하기 위해서는 다른 접근법이 필요한 것으로 분석된다.

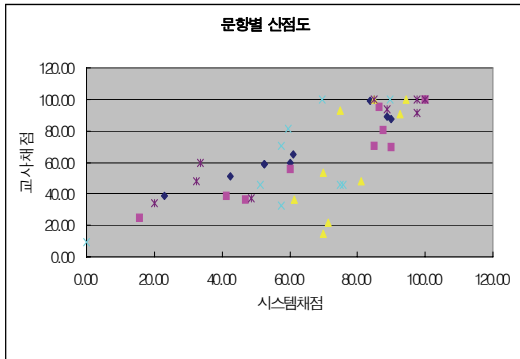


그림 3. 문항별 산점도

각 문항에 대한 산점도를 그림 3에 표시하였다. 파란 ◇는 1번 문항, 노란 △는 2번 문항, 고동색 *는 3번 문항, 하늘색 ×는 4번 문항, 주황색 □는 5번 문항을 나타낸다. 1번 문항과 3번 문항, 5번 문항은 일차함수처럼 비례하는 분포를 보인다. 이것은 상관계수가 1에 가까운 것을 보여주고 있어 교사채점과 시스템 채점의 일치도가 높음을 말하고 있다. 실제 상관계수도 [그림 2]에서 알 수 있듯이 높게 나타나 있다. 그러나 문항 2와 문항 4의 경우 그 분포가 직선에서 떨어지는 형태를 보여 상관계수가 낮음을 알 수 있다. 순서를 나타내는 문항 2와 장단점을 제시하는 문항 4의 경우는 방법의 개선이 필요함을 알 수 있다.

5. 결론 및 향후 과제

서답형 평가에 대한 연구의 필요성에 따라 최근 많은 연구가 이루어지고 있다. 그렇지만 그 연구의 방법이 일반화된 방법을 사용하여 질의 유형의 특수성을 고려하지 못하고 있다. 본 연구는 이 문제를 해결하고자 질의 유형에 따라 적응 처리하는 서답형 자동채점 시스템을 제안하고 실험하였다. 그 결과 5개 유형 가운데 4개 유형에 대해 제안한 시스템이 더 나은 결과를 보였다. 나머지 1개와 4개 가운데 성능이 저조한 1개의 질의 유형에 대한 연구는 앞으로 개선이 필요하다.

순서제시형의 유형에서 영역을 구분하는 방법과 순서 정보를 반영하는 방법에 대한 연구가 필요하다. 그리고 장단점 제시형에서 장점과 단점을 혼재하여 기술할 경우 이를 구분할 수 있는 연구가 필요하다. 그리고 서답형 질의 유형에 대한 연구를 통해 질의 유형의 확장이 필요하다. 또한 각 질의 유형에 대해 본 연구는 기초적인 자연어처리기술을 적용하였지만 앞으로 구문해석, 의미해석 등의 고급 자연어처리 기술을 이용하여 성능을 개선할 필요가 있다.

참 고 문 헌

- [1] 진경애, 이병천, 신동광, 박태준, 주현우, 서답형 문장 자동채점 프로그램 도입방안 연구(III), 한국교육과정평가원 연구보고서, 2008.
- [2] 성태제, 양길석, 강태훈, 정은영, 학업성취도 평가 서답형 문항 컴퓨터채점화 방안 탐색, 한국교육과정평가원 연구보고서, 2010.
- [3] 박정, “서답형 평가 문항 자동채점프로그램 활용가능성 탐색,” 교육평가연구, 제22권, 제3호, pp.615-631, 2009.
- [4] Donald E. Powers, Jill C. Burstein, Martin Chodorow, Mary E. Fowles, and Karen Kukich, “Stumping e-rater:challenging the validity of automated essay scoring,” Computers in Human Behavior, Vol.18, pp.103-134, 2002.

[5] Salvatore Valenti, Francesca Neri, and Alessandro Cucchiarelli, "An Overview of Current Research on Automated Essay Grading," J. of Information Technology Education, Vol.2, pp.319-330, 2003.

[6] Rehab M. Duwairi, "A framework for the computerized assessment of university student essays," Computers in Human Behavior, Vol.22, pp.381-388, 2006.

[7] 배화식, 정순호, "자연어처리과정을 이용한 웹기반 한문 주관식 채점 시스템," 2005년 한국정보처리학회 학술대회논문집, 제12권, 제2호, pp.813-816, 2005.

[8] 조우진, 오정석, 이재영, 김유섭, "의미커널과 한글 워드넷에 기반한 지능형 채점 시스템," 정보처리학회논문지 A, 제12-A권, 제6호, pp.539-546, 2005.

[9] 김용범, 김유섭, "한국어 워드넷에서의 개념유사도를 활용한 선택형 문항 생성시스템," 정보처리학회논문지 A : 제15-A권, 제2호, pp.125-134, 2008.

[10] 조지민, 김경훈, "서답형 문항의 인터넷 기반 채점 시스템 설계 연구," 한국컴퓨터교육학회논문지, 제10권, 제2호, pp.89-100, 2007.

[11] 황도삼, 최기선, 김태석 공역, 자연언어처리, 홍릉과학출판사, 2007.

[12] 김명철, 김덕봉, 이하규, 김유성, 김재훈, 박혁로 공역, 최신정보검색론, 홍릉과학출판사, 2001.

[13] 방훈, 강태호, 김원진, 원대희, 이재영, "웹기반 주관식 답분류 채점 시스템," 2001년 한국정보과학회 학술대회논문집, 제28권, 제2호, pp.589-591, 2001.

[14] 송미영, 김수진, 김희경, 남명호, "온라인시스템을 활용한 대규모서답형 평가의 채점 일관성," 교육평가연구, 제22권, 제3호, pp.827-846, 2009.

[15] 정은미, 프로그래밍언어에 대한 주관식 자동채점 시스템의 설계 및 구현, 안동대학교 대학원 석사논문, 2008.

[16] 김흥기, 홍동권, "주관식 평가를 위한 웹기반 온라인 평가시스템의 구현 및 적용," 한국정보교육학회 논문지, 제8권, 제2호, pp.251-260, 2004.

[17] 박희정, 강원석, "유의어 사전을 이용한 주관식 문

제 채점 시스템 설계 및 구현", 한국컴퓨터교육학회 논문지, 제6권, 제3호, pp.207-216, 2003.

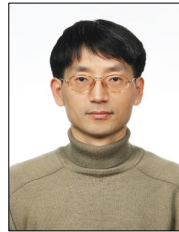
[18] 강원석, "구문-격의미 정보를 이용한 주관식 문제 채점 시스템 설계 및 구현", 한국컴퓨터교육학회 논문지, 제10권, 제5호, pp.61-69, 2007.

[19] 김재훈 외 5인, "새 환경에 적용가능한 한국어 품사 태깅 시스템 KTAG99," 11회 한글 및 한국어정보처리 학술대회논문집, 1999.

저 자 소 개

강 원 석(Won-Seog Kang)

정회원



- 1985년 2월 : 경북대학교 전자공학과(공학사)
 - 1988년 2월 : 한국과학기술원 전산학과(공학석사)
 - 1995년 2월 : 한국과학기술원 전산학과(공학박사)
 - 1995년 3월 ~ 현재 : 안동대학교 정보과학교육과 교수
- <관심분야> : 자연어처리, 정보검색