

# 구문의미트리 비교기를 이용한 주관식 문항 채점 시스템에 대한 연구

강원석<sup>†</sup>

## 요 약

주관식 문항은 깊이 있는 사고능력 평가와 고등정신능력 평가에 적합하나 채점하기가 쉽지 않다. 동일한 채점기준을 갖더라도 채점자에 따라 다른 채점결과를 가져올 수 있으므로 객관적인 자동 채점 시스템이 필요하다. 그렇지만 채점 시스템은 표현 언어인 한국어 분석과 비교의 문제가 걸려있다. 본 연구는 한국어의 구문 분석을 구현하고 결과인 구문분석 트리를 비교하는 비교기를 이용하여 주관식 문항을 채점하는 채점 시스템을 설계, 개발하였다. 이 시스템은 단어 중심의 채점 시스템과 구문의미트리 중심의 채점 시스템을 복합한 시스템으로 구문의미트리 비교기를 활용하였다. 본 시스템의 실험 결과 제안한 구문의미트리 중심의 채점 시스템과 복합 채점 시스템이 더 나은 결과를 가져옴을 알 수 있었다. 본 연구는 한국어 구문의미분석과 주관식 채점 영역에 활용할 수 있고 또한 문서 분류에도 활용할 수 있을 것이다.

주제어 : 주관식 채점, 구문의미트리, 비교기, 유사도

## Research on Subjective-type Grading System Using Syntactic-Semantic Tree Comparator

WonSeog Kang<sup>†</sup>

### ABSTRACT

The subjective question is appropriate for evaluation of deep thinking, but it is not easy to score. Since, regardless of same scoring criterion, the graders are able to produce different scores, we need the objective automatic evaluation system. However, the system has the problem of Korean analysis and comparison. This paper suggests the Korean syntactic analysis and subjective grading system using the syntactic-semantic tree comparator. This system is the hybrid grading system of word based and syntactic-semantic tree based grading. This system grades the answers on the subjective question using the syntactic-semantic comparator. This proposed system has the good result. This system will be utilized in Korean syntactic-semantic analysis, subjective question grading, and document classification.

**Keywords** : Subjective-type Grading, Syntactic-Semantic Tree, Comparator, Similarity

<sup>†</sup> 정 회 원: 안동대학교 교수(교신저자)

논문접수: 2018년 7월 31일, 심사완료: 2018년 10월 15일, 게재확정: 2018년 10월 16일

\* 본 논문은 2016년 안동대학교 연구비에 의하여 연구되었음.

## 1. 서론

주관식 문항은 깊이 있는 사고능력 평가와 고등정신능력 평가에 적합하여 필요성이 증가하고 있다. 그렇지만 학교 현장에서는 채점기준을 선정하는 것도 쉽지 않고 동일한 채점기준을 적용하더라도 채점자에 따라 다른 결과를 가져올 수 있어 어려움이 있다. 최근 이를 해결하려는 자동 채점 시스템에 대한 연구가 이루어지고 있다. 이 연구들은 서술된 답변의 한국어 문장을 형태소 해석하여 단어를 추출하고 단어들의 출현빈도 등을 근거로 자동채점을 시도하였다[1][2].

조우진과 동료들(2005)의 연구는 형태소 해석의 한계를 극복하기 위하여 단어의 개념 정보를 추출하고 이를 채점에 반영하였다[3]. 그렇지만 이 방법은 단어간의 구문 관계, 의미관계는 반영하지 못하고 있다.

강원석(2007)의 연구는 답변의 문장에 대한 구문의미분석을 적용하고 그 결과인 구문, 격의미 정보를 활용하나 구문의미분석의 결과인 구문의미트리를 비교하는 방식에 한계가 있다[4].

강원석(2017)의 연구는 구문의미분석의 결과인 구문의미트리를 비교하는 비교기를 제안하였다[5]. 그렇지만 이를 어떻게 주관식 채점에 반영할지에 대한 것은 논의하지 않았다.

본 연구는 구문의미트리 비교기를 활용하여 서술식 답변 채점의 성능을 향상하는 복합 채점 시스템을 설계, 구현한다. 이 시스템은 단어간의 구문, 의미관계를 분석하는 구문의미해석기, 구문의미해석의 결과인 구문의미트리를 비교하는 구문의미트리 비교기, 그리고 단어 중심의 채점과 구문의미트리 중심의 채점을 복합하는 복합채점기 등을 이용하여 주관식 문항을 채점한다.

2장은 주관식 문항 채점시스템의 관련 연구를 기술하고 3장은 구문의미트리 비교기를 이용한 복합 채점 시스템을 기술한다. 4장은 시스템의 실험과 결과 분석을 기술하고 5장은 결론과 논의사항을 기술한다.

## 2. 관련 연구

주관식 문항 채점 시스템과 관련되는 연구는

크게 채점 보조 시스템, 영어 논술 채점, 어휘 중심의 채점, 기계학습 중심의 채점, 구문트리 중심의 채점의 연구로 구분할 수 있다.

### 2.1 채점 보조 시스템 연구

김동기와 홍동권(2004)은 주관식 평가를 지원하기 위해 웹기반의 온라인 채점 보조 시스템을 제안하였다[6]. 이재영(2011)은 채점을 위해 배점 기준표에 없는 유형의 경우 기준표 작성자를 통해 배점 기준표를 받아 공정한 채점을 할 수 있는 관리 시스템을 제안하였다[7]. 조지민과 김경훈(2014)은 서답형 문장 채점자의 일관성 있는 채점 유지를 위한 온라인 채점 시스템을 구축하고 신뢰성을 평가하였다[8]. 천민아와 동료들(2015)은 빈도수 높은 학생답안을 모범답안으로 하여 채점자가 채점을 하고 이를 초기 모델로 학생답안의 유형을 늘려가는 준지도학습 방식의 채점시스템을 제안하였다[9]. 이 연구들은 채점자의 채점을 보조하고 도와주는 관리 시스템에 대한 것으로 시스템이 채점하는 자동채점에 대한 것은 아니다.

### 2.2 영어 논술 중심의 채점 연구

Claudia와 Martin(2003)은 영어문장에 대한 통사적 변화, 대명사 교환, 형태론적 변환, 유의어 치환 등의 과정을 통해 표준표현으로 변환한 후 모범 답안과 비교하여 채점을 하는 연구를 하였다[10]. 이 과정에 구문 분석과 유의어 활용의 과정이 적용되었다. Attali와 Burstein(2006)은 영어 논술의 자동채점을 위해 전문가의 채점 기준 세트를 근거로 학생들의 답안에 대해 답화구조, 어휘 사용, 통사적 변수 등을 검사하였다[11]. 이 연구들은 어휘분석과 구문분석을 하고 있지만 그 기술들은 영어에 대한 것이므로 어휘와 문법구조가 다른 한국어에 적용할 수가 없다.

### 2.3 어휘 중심의 채점 연구

박희정과 강원석(2003)은 단어간의 비교를 통한 채점의 한계를 극복하기 위해 유의어 사전을 활용하여 유사한 의미를 가진 단어간의 의미적 비교를 가능하게 하였다[13]. 정상목과 동료들(2005)

은 대량의 데이터에 든 의미구조를 찾아내는 LSA(Latent Semantic Analysis) 분석을 이용하여 채점을 하였다[1]. 조우진과 동료들(2005)은 대량의 말뭉치를 이용하여 의미커널을 구축하고 답안을 채점할 때 구축된 의미커널과 한글 워드넷을 이용하여 답안을 채점하였다[3]. 이 연구들은 형태소 분석의 결과인 단어를 추출하고 대량의 데이터 속에 은닉된 단어간의 의미관계나 사전을 통한 단어의 의미적 유사어 추출 기능을 이용하여 채점을 시도하였다. 그렇지만 단어간 거리나 구조적 관계를 고려한 구문분석이 이루어지지 않아 채점에 한계가 있다.

조정현과 동료들(2009)은 학생답안의 내용이 여러 개의 정답 집합과 얼마나 일치하는가의 여부에 따른 채점을 하였다[2]. 이 연구도 단어를 추출한 후 단어집합의 매치 여부만 비교하여 단어간의 구문 관계 등을 이용하지 못하고 있다.

강원석(2011)은 주관식 문제 유형을 정의하고 그 유형에 따른 채점 방식을 다르게 한 방법을 적용하였다[14]. 이 연구도 단어 중심의 비교를 통한 채점방식을 적용하여 구문적 관계를 이용하지 못하였다. 박일남과 동료들(2013)은 단어와 구 수준의 정답 템플릿을 설계하고 이를 토대로 답안과 일치여부를 검사하였다[15]. 이 연구도 유의어를 활용하여 개념 중심의 일치여부를 검사하나 구문적 관계를 이용하지 못하고 있다. 강승식과 장은서(2016)은 학생답안을 분석하여 정답 템플릿을 생성한 후 채점을 하였다[16]. 이 연구는 개념과 실마리 단어 등을 활용하여 채점의 질을 향상하였으나 단어 구 수준의 채점으로 한계가 있다.

## 2.4 기계학습 중심의 채점 연구

허정만과 박소영(2013)은 여러 종류의 기계학습 분류기를 이용하여 답안이 가지고 있는 텍스트 구성비, 키워드 비, 부사와 숫자의 조합자질을 검사하여 만장일치체의 채점을 적용하였다[17]. 천민아와 동료들(2016)의 연구도 여러 개의 기계학습 분류기를 적용하여 만장일치의 경우를 제시하는 채점 보조 시스템을 제안하였다[18]. 이 연구들은 자동학습의 분류기를 활용하고 있다. 그렇지만 자동학습의 입력으로 단어집합을 사용하였기 때문

에 단어간의 거리적 관계나 구문관계 등은 반영되지 못하였다.

## 2.5 구문트리 중심의 채점 연구

강원석(2007)은 구문의미분석을 통해 구문의미 정보를 추출하고 이를 활용한 채점을 시도하였다[4]. 그렇지만 구문의미분석의 결과인 구문의미트리를 비교할 때 단지 일치여부만 검사하여 유사한 구문의미트리를 구분하지 못하였다. 강원석(2017)은 기존 연구의 한계를 해결하기 위해 구문의미트리를 비교하는 비교기를 제안하였다[5]. 이 비교기는 두 쌍의 구문의미트리에 대해 유사한 정도를 계산하는 도구로 개발하여 좋은 성능을 얻었다. 그렇지만 답변의 문장들이 여러 개의 구문의미트리들로 구성되므로 구문의미트리 비교기를 그대로 적용할 수 없다. 본 연구는 여러 개의 구문의미트리로 이루어지는 답변의 채점을 위해 구문의미트리 비교기를 어떻게 적용할 것인지를 연구하였다.

## 2. 구문의미트리 비교기를 이용한 주관식 채점 시스템

### 3.1 구문의미트리 비교기

구문의미트리 비교기는 구문의미트리 간의 유사성을 비교하기 위해 다음과 같은 원칙을 적용하였다.

원칙 1) 두 구문의미트리의 유사성은 두 트리의 루트의 유사성과 서브트리 유사성의 합이다.

구문의미트리는 하나의 루트노드와 여러 개의 서브트리로 이루어진다. 두 구문의미트리가 같은지는 루트노드가 서로 같은지 그리고 여러 개의 서브트리 가운데 서로 같은 것이 있는지를 비교한다. 따라서 전체의 유사도는 루트노드의 유사성과 서브트리의 유사성의 합으로 정의한다.

원칙 2) 루트의 유사성은 서브트리 유사성보다 더 중요하다.

구문의미트리에서 루트는 동사로서 술어의 유형과 의미를 결정하고 서브트리는 그 유형에 따른 역할(격)을 채우는 것이므로 역할이 서로 다르

다. 따라서 술어의 유형을 결정하는 루트의 유사성을 서브트리의 유사성보다 더 가중한다.

원칙 3) 루트의 유사성은 노드를 구성하는 어휘의 의미 유사성으로 정의한다.

루트노드를 이루는 단어는 트리의 유형을 결정하는 정보를 담고 있다. 따라서 루트노드의 유사성을 비교하기 위해 단어가 가지고 있는 유형과 의미 정보를 추출하는 의미속성 추출기를 이용하고 추출된 의미속성의 유사도 함수를 정의하여 이용한다.

원칙 4) 서브트리의 유사성은 격이 같은 서브트리의 유사성의 합이다.

루트노드는 여러 개의 자식 서브트리를 가진다. 각 자식 서브트리는 서로 역할(격)이 다르다. 따라서 같은 역할을 하는 서브트리끼리 비교하여 그 유사성을 도출하고 그 결과를 합한다.

원칙 5) 서브트리의 유사성은 서브트리 수에 의존하고 한계치를 넘지 못한다.

서브트리의 유사성은 같은 역할을 하는 서브트리의 유사성의 합이라고 정의하지만 서브트리 수가 많을 경우와 적을 경우 반영하는 가중치가 다르다. 즉, 서브트리 수가 많을 경우 더 가중하고 그 가중치는 한계치를 넘지 않도록 정의한다.

위와 같은 5개의 원칙을 고려하여 구문의미트리 비교 시스템에 사용할 유사도 계산식을 다음과 같이 정의하였다.

$$ST(T_1, T_2) = \frac{SL(r(T_1), r(T_2)) \times \alpha + \frac{\sum_{(T_{1i}, T_{2j}) \in T_1 \cap T_2} ST(T_{1i}, T_{2j}) \times 2}{n} \times \beta \times \gamma}{\alpha + \beta \times \gamma} \quad (식1)$$

$ST(T_1, T_2)$ 는 구문의미트리  $T_1$ 과  $T_2$ 의 유사도를 계산하는 함수이다.  $ST(T_1, T_2)$ 는 원칙 1에서 제시한 것과 같이 루트의 유사성과 서브트리의 유사성의 합으로 구성된다. 그리고 원칙 2에서 루트노드의 가중치는  $\alpha$ 로 표현하고 서브트리의 가중치는  $\beta$ 로 표현한다.

$SL(r(T_1), r(T_2))$ 는  $T_1$ 의 루트와  $T_2$ 의 루트의 어휘 의미 유사성을 계산하는 함수로 다음과 같이 정의된다.

$$SL(a, b) = \frac{2 * |sem(a) \cap sem(b)|}{|sem(a)| + |sem(b)|} \quad (식2)$$

이 함수는 다이스 계수를 이용하여 정의하였다.  $sem(a)$ 는 단어  $a$ 의 의미속성을 추출하여 얻은 의미속성집합을 나타내고  $|sem(a)|$ 는 의미속성집합의 원소수를 나타낸다.

원칙 4에 의하면 서브트리의 유사성은 격이 같은 서브트리의 유사성의 합으로 정의된다. 식(1)의  $\Sigma$ 가 그 합을 표현하고  $\Sigma$  표기의 조건에 있는 교집합은 격이 같은 서브트리를 의미한다. 즉,  $(T_{1i}, T_{2j}) \in T_1 \cap T_2$ 는  $T_1$  트리의  $i$ 번째 서브트리가  $T_2$  트리의  $j$ 번째 서브트리와 격이 같음을 나타낸다. 이때 격이 같은 트리는 양 구문의미트리에 모두 있으므로 2로 곱하여 가중한 후 정규화를 위해 전체 서브트리 수  $n$ 으로 나눈다. 최종적으로 계산된 값에 서브트리 유사성 가중치  $\beta$ 를 곱한다.

$$n = |T_1 \cap T_2| \times 2 + |T_1 - T_2| + |T_2 - T_1| \quad (식3)$$

$n$ 의 계산에 사용한  $|T|$  기호는 트리  $T$ 의 서브트리수를 의미한다.  $n$ 은 비교할 두 구문의미트리에서 전체 서브트리수를 나타내는 것으로 격이 같은 공통 서브트리수의 2배에  $T_1$ 에만 있는 서브트리수와  $T_2$ 에만 있는 서브트리수를 더한 값이다.

$$\gamma = (1 - \frac{1}{e^{nk}}) \quad (식4)$$

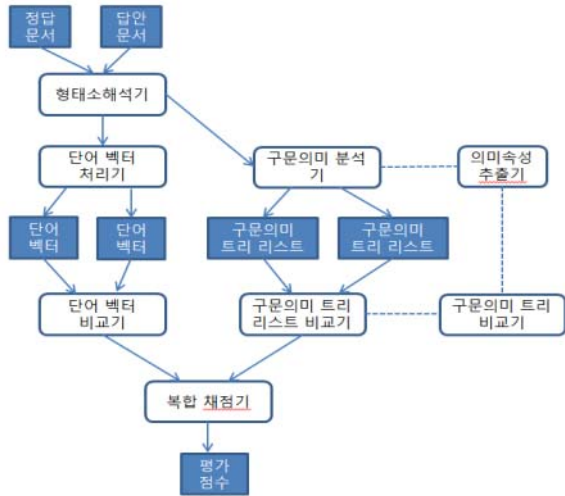
가중치  $\gamma$ 는 원칙 5를 반영한 것으로 서브트리의 수가 증가함에 따라 가중치가 증가하나 한계값까지만 증가하도록 설계하였다.  $k$ 는 감마값을 결정짓는 감마계수상수로 실험을 통해 정의하였다.

### 3.2 구문의미트리 비교기를 이용한 주관식 복합 채점 시스템

주관식 문항 채점을 위한 시스템은 [그림 1]과 같이 구성하였다.

전체 시스템은 형태소 해석기, 단어벡터 처리기, 단어벡터 비교기, 구문의미 분석기, 의미속성 추출기, 구문의미트리 비교기, 구문의미트리 리스트 비교기, 복합 채점기로 구성된다.

형태소 해석기는 [19]를 이용하였다. 해석결과는 단어벡터 처리기와 구문의미 분석기로 전달된다.



[그림 1] 주관식 문항 채점 시스템

단어벡터 처리기는 형태소와 태그로 구성된 형태소 해석의 결과에서 독립소인 단어를 추출하고 이를 (단어, 가중치)의 리스트인 벡터로 변환한다. 단어벡터 비교기는 변환된 두 벡터간의 유사성을 비교하는 비교기이다. 유사성의 계산식은 코사인 계수를 이용한 것으로 (식 5)와 같다. WA, WB는 정답 A, 학생답안 B에 대한 단어벡터이다.

$$wsim(WA, WB) = \frac{\sum_{i=1}^n WA_i \times WB_i}{\sqrt{\sum_{i=1}^n WA_i^2} \times \sqrt{\sum_{i=1}^n WB_i^2}} \quad (식5)$$

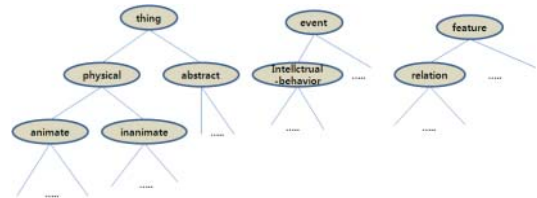
구문의미 분석은 [5]를 이용하였다. 형태소 해석과 구문의미 분석기의 수행 예는 표 1과 같다.

<표 1> 형태소해석과 구문의미분석 예

| 항목                  | 내용  |
|---------------------|---|
| 답안문서                | 소스의 가독성이 높아진다.<br>...                             |
| 답안문서<br>형태소해<br>석결과 | 소스/ncn+jcm, 가독성/ncn+jcs, 높아지<br>/pvg+ef+sf<br>... |
| 구문의미<br>분석기의<br>결과  | ( FINEN 높아지 ( SUBJ 가독성 ( ADNM 소스 ) ) )<br>...     |

구문의미 분석기는 구문의미분석에 필요한 제한 조건 등의 검사를 위해 단어에 대한 의미속성을 추출하는 의미속성 추출기를 사용한다. 의미속성 추출기는 단어에 대한 의미의 상하위 의미를 추출하는 것으로 [20]의 시소러스 사전을 이용하였다. 시소러스 사전은 최상위로 thing, event, feature의 개념을 가지는 트리로 구성된다. 그 일

부는 [그림 2]와 같다. 예 단어 ‘소스’는 크게 thing의 범주에 포함되고 그 하위 개념을 나열하면 ‘domain edible-thing food inanimate-thing industry material physical-thing thing 식료품’과 같다.



[그림 2] 상하위 의미 체계의 일부

구문의미 분석기는 답안에 대한 구문의미트리들을 생성해낸다. 한 답안에 대한 여러 개의 구문의미트리들을 리스트로 정의하면 답안과 정답에 대한 구문의미트리 리스트 TA와 TB는 다음과 같다.

$$TA = [TA_1, TA_2, TA_3, \dots, TA_m]$$

$$TB = [TB_1, TB_2, TB_3, \dots, TB_n]$$

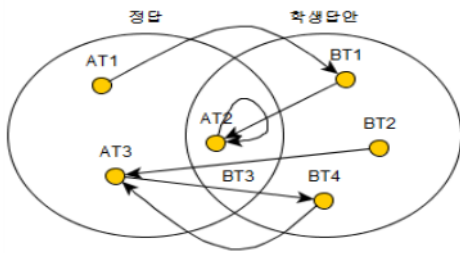
구문의미트리 비교기는 한 쌍의 구문의미트리 리스트를 비교하는 것이 아니라 한 쌍의 구문의미트리를 비교한다. 따라서 리스트를 비교하기 위해서는 구문의미트리 리스트 비교기를 사용해야 한다.

단어 중심의 채점 시스템은 단어를 추출하고 이를 벡터로 만들어 비교하는 방식으로 정답과의 유사도를 계산하여 채점한다. 구문의미트리 중심의 채점 시스템도 구문의미트리를 추출하고 이를 벡터로 만들어 비교하는 방식으로 정답과의 유사도를 계산할 수 있다. 단어 벡터의 경우 같은 단어인지 아닌지에 따라 벡터의 축이 같은 것인지 아닌지 구분하여 비교한다. 그러나 구문의미트리인 경우 정확히 일치하지 않지만 유사한 구문의미트리인지 아닌지를 식별하기가 쉽지 않다. 강원석(2015)의 연구에서는 한계값 이상의 유사한 구문의미트리를 찾아내고 이를 같은 벡터의 축으로 인식하도록 같은 구문의미트리로 복사한 후 코사인 계수와 같은 유사도 계산방식을 적용하였다 [21]. 본 연구에서는 한계값을 사용하지 않고 구문의미트리 유사도 값을 그대로 계산에 적용한 구문의미트리 리스트 비교기를 구성하였다. 구문의미트리 리스트 비교기에 적용한 원칙과 식은 다

음과 같다.

원칙 6) 정답 리스트의 각 구문의미트리에 대해 학생 답안 리스트에서 가장 유사한 구문의미트리를 찾고 그 유사성을 합한다.

이 원칙은 정답의 구문의미트리와 유사한 것이 학생답안에 얼마나 있는지 그 정도를 나타내기 위한 것이다.



[그림 5] 구문의미트리 리스트 비교

그림을 들어 설명한다. [그림 4]의 원은 정답과 학생답안의 구문의미트리 리스트를 나타낸다. 정답의 리스트는 AT1, AT2, AT3의 구문의미트리로 구성되고 학생답안은 BT1, BT2, BT3, BT4의 구문의미트리로 구성된다고 하자. AT1과 가장 유사한 것 BT1을 화살표로 표현하였다. 그러면 AT1뿐 아니라 정답의 구문의미트리 모두에 대해 가장 유사한 것을 찾고 그 유사도를 합하여 유사성을 계산한다. 즉  $\sum_{i=1}^m cal(AT_i, BT)$ 을 계산한다.  $cal(AT_i, BT)$ 는 구문의미트리 리스트 BT에서 AT<sub>i</sub>와 가장 유사한 구문의미트리를 찾아 그 유사한 정도를 계산하는 함수이다.

원칙 7) 학생 답안 리스트의 각 구문의미트에 대해 정답 리스트에서 가장 유사한 구문의미트리를 찾고 그 유사성을 합한다.

두 리스트간의 유사성을 나타내기 위해 한 리스트 관점에서만 볼 것이 아니라 양방에서 보고 합산해야 한다. 이 원칙은 반대 관점인 학생답안의 구문의미트리와 유사한 것이 정답에 얼마나 있는지 그 정도를 나타내기 위한 것이다.

원칙 8) 구문의미트리 리스트 유사도는 정답의 구문의미트리들에 대한 유사도의 합과 학생답안의 구문의미트리들에 대한 유사도의 합을 전체 구문의미트리 수로 나눈다.

양방의 관점에서 유사한 정도를 계산한 후 이를 평균하기 위한 원칙이다. 이 세 원칙을 반영한

계산식은 다음과 같다.

$$lsim(TA, TB) = \frac{\sum_{i=1}^m cal(TA_i, TB) + \sum_{j=1}^n cal(TB_j, TA)}{m+n} \quad (식6)$$

$cal(TA_i, TB)$ 은 TB의 구문의미트리들 가운데 구문의미트리 TA<sub>i</sub>와 가장 유사한 것을 찾고 그 유사도값을 계산한다. 그리고 반대방향의 관점에서 유사도 합을 계산한 후 전체 구문의미트리 수로 나누어 평균을 구한다.

마지막 복합채점기는 단어 중심의 채점 시스템과 구문의미트리 중심의 채점시스템을 복합한 것이다. 학생 답안의 유형에 따라 단어 중심의 채점이 더 좋은 결과를 가져오는 것도 있고 구문의미트리 중심의 시스템이 더 좋은 결과를 가져오는 것도 있을 것이다. 본 연구에서는 어떤 유형이라도 좋은 채점결과를 얻도록 두 채점 시스템을 복합하였다.

$$sim(A, B) = \frac{\mu \times wsim(WA, WB) + \nu \times lsim(TA, TB)}{\mu + \nu} \quad (식7)$$

식7의 A와 B는 정답과 학생답안이고 WA, WB는 형태소 해석의 결과로 나온 단어벡터이고 TA, TB는 구문의미분석의 결과로 나온 구문의미트리 리스트이다.  $\mu$ 와  $\nu$ 는 단어중심의 채점 시스템과 구문의미트리 중심의 채점시스템에 대한 가중치이다.

### 3. 실험 및 결과 분석

본 연구에서 제안한 주관식 채점 시스템의 실험을 위해 컴퓨터 분야의 전공과목 시험에서 33쌍의 검사 답안을 수집하였다.

<표 2> 검사답안의 예

| 학생id | 학생 답안  |
|------|--|
| 01   | 열거형이 제공되지 않으면 프로그래머는 리터럴 값에 정수값을 대응시켜 표현해야 된다. 이것은 프로그램을 읽고 이해하기 어렵게 한다. 열거형을 정의하는 실비는 쉬운 방법으로 프로그래밍 언어의 능력을 크게 증가시켜 주는 효율적인 기법이다. |
| 02   | 열거형 자료를 통해 정수값만을 사용해서 표현해야 할 것들이 손쉽게 표현된다. 프로그램의 판독성을 증가한다. 기억영역을 추가로 필요로 하지 않는다.  |
| ...  | ...  |

검사 답안은 평균 33단어와 255바이트로 구성되어 있다. 검사 답안에 대해 사람이 채점한 값의

분포는 0.8 이상은 34%, 0.4이상 0.8미만은 42%, 0.4미만은 24%이다.

채점 시스템의 성능은 사람이 채점한 점수와 시스템이 채점한 점수의 상관계수로 측정하였다. 그 식은 다음과 같다.

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_X \sigma_Y} \quad (\text{식8})$$

X는 사람이 채점한 점수이고 Y는 시스템이 채점한 점수이다.  $\bar{X}$ 와  $\bar{Y}$ 는 사람과 시스템이 채점한 점수의 평균이고  $\sigma_X$ 와  $\sigma_Y$ 는 사람과 시스템이 채점한 점수의 표준편차이다. 그리고 n은 학생답안의 개수이다.

본 연구에서는 두 가지 유형의 시스템에 대한 실험을 하였다.

- 실험1 : 구문의미트리 비교기를 이용한 구문의미트리 중심의 채점 시스템의 실험
- 실험2 : 단어 중심의 채점과 구문의미트리 중심의 채점의 복합 시스템의 실험

#### 4.1 구문의미트리 비교기를 이용한 구문의미트리 중심의 채점 시스템의 실험

구문의미트리 비교기의 (식 1)에 반영되는  $\alpha$ 와  $\gamma$ 에 대해 여러 가지 패턴에 대한 구문의미트리 비교기를 사용하였을 때 학생답안에 대한 채점의 실험 결과는 <표 3>와 같다.

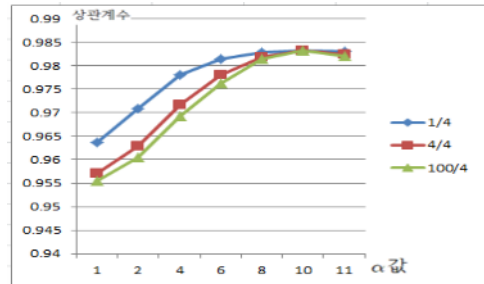
<표 3>  $\alpha$ 값과  $\gamma$ 계수값에 따른 채점시스템 실험 결과

| $\alpha$ \ k | 1/4      | 4/4      | 100/4    |
|--------------|----------|----------|----------|
| 1            | 0.963732 | 0.957018 | 0.955444 |
| 2            | 0.970973 | 0.962924 | 0.960429 |
| 4            | 0.978092 | 0.971753 | 0.969254 |
| 6            | 0.981433 | 0.978005 | 0.976296 |
| 8            | 0.982843 | 0.981848 | 0.981393 |
| 10           | 0.983336 | 0.983336 | 0.983336 |
| 11           | 0.983066 | 0.982537 | 0.982094 |

표의 행은  $\alpha$ 값을 나타내고 열로는  $\gamma$ 계산 ( $\gamma = (1 - \frac{1}{e^{nk}})$ )에 사용되는  $\gamma$ 계수 k값을 나타낸다. 이때 서브트리의 가중치  $\beta$ 는 1로 고정하였다. 표의 행과 열에 해당하는 값으로 21개의 구문의미트리 채점 시스템을 구축하고 상관계수 값으로 각 채점시스템의 성능을 구하였다.

#### - 분석1

실험 결과에서 먼저  $\alpha$ 값의 변화가 채점에 어떠한 영향을 미치는지를 알아보기 위해 <표 3>의 내용을 그래프로 도시하였다.

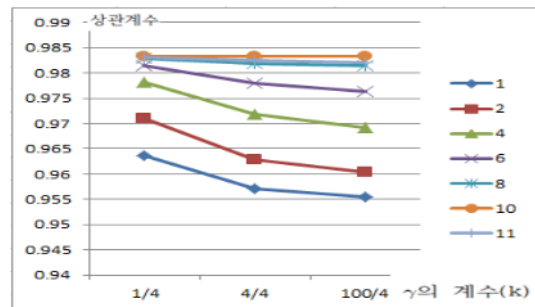


[그림 6]  $\alpha$ 값의 변화에 따른 채점성능

그래프에 도시된 것과 같이 검사데이터에 대해  $\alpha$ 값이 10일 때 가장 채점성능이 좋은 것을 알 수 있다. 이것은 두 구문의미트리의 유사성이 루트와 서브트리의 유사성의 합으로 구성된다의 원칙 1과 루트가 서브트리보다 가중된다는 원칙 2가 유효함을 알 수 있었다.

#### - 분석 2

$\gamma$ 에 영향을 미치는  $\gamma$ 계수 값의 변화에 따른 채점의 성능을 그래프로 표현한 것은 [그림 5]와 같다.



[그림 7]  $\gamma$ 계수값의 변화에 따른 채점 성능

$\gamma$ 는 서브트리의 수에 따른 가중치 값을 의미한다. 서브트리 수가 많으면 더 가중되도록 하였다.

$\gamma = (1 - \frac{1}{e^{nk}})$  이므로 서브트리 수 n이 클수록 감마 값이 1에 가깝게 되고 n이 작으면 0에 이른다. 실제로 서브트리가 하나도 없으면 감마는 0이 된다. 식에서  $\gamma$ 계수 값이 100/4인 경우 n에 관계없이  $\gamma$  값이 1에 가깝게 된다. 이 사실은 원칙5를 적용하지 않은 결과이다. 즉, 서브트리 수에 따른 가중치를 반영하지 않은 결과이다. 그래프에서 보여주



듯이  $\gamma$  계수값이 1/4일 때 더 좋은 성능을 보여주고 있으므로 원칙 5가 타당함을 입증한다.

$\alpha$ 값과  $\gamma$  계수값에 따른 21개 시스템 가운데 성능이 가장 좋은 것은  $\alpha$  값이 10이고  $\gamma$  계수값이 1/4인 시스템이다. 이 시스템의 상관계수는 0.983336으로 단어중심의 채점 시스템의 상관계수 0.96869보다 더 좋다. 이것은 구문의미트리 중심의 채점이 타당성 있음을 보여주고 있다.

#### 4.2 복합 채점 시스템에 대한 실험

기존의 많은 주관식 문항 채점시스템들이 단어 중심의 채점 방식을 적용하였다. 본 연구에서는 구문의미트리 중심의 채점방식을 제안하였다. 실험 1에서 설명한 것과 같이 단어 중심의 채점 시스템의 결과 0.96869보다 더 좋은 0.983336의 상관계수 값을 얻었다. 그렇지만 학생답안 유형에 따라 단어중심의 채점이 더 좋은 것이 있을 수도 있으므로 두 방식을 복합하여 시스템을 구성하였다. 최적의 복합채점 시스템을 찾기 위해 (식 7)의  $\mu$ 값과  $\nu$ 값을 달리하여 시스템을 구성하고 각 시스템의 실험 결과를 <표 4>에 기술하였다.

<표 4> 복합채점 시스템 실험 결과

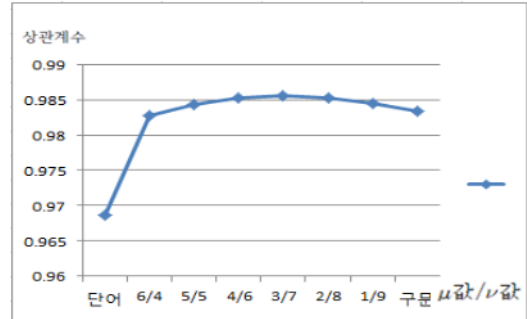
| 시스템종류       | $\mu$ | $\nu$ | 상관계수     |
|-------------|-------|-------|----------|
| 단어중심채점      | 10    | 0     | 0.96869  |
| 구문의미트리 중심채점 | 0     | 10    | 0.983336 |
| 복합채점        | 6     | 4     | 0.982747 |
|             | 5     | 5     | 0.984303 |
|             | 4     | 6     | 0.985199 |
|             | 3     | 7     | 0.985495 |
|             | 2     | 8     | 0.985248 |
|             | 1     | 9     | 0.984511 |

$\mu$ 값은 단어중심의 채점 시스템에 대한 가중치이고  $\nu$ 값은 구문의미트리 중심의 채점 시스템에 대한 가중치이다(구문의미트리 중심의 채점 시스템에 사용한 식 1의  $\alpha$ 값과  $\gamma$ 계수 상수는 10과 1/4이다).  $\mu$ 값이 10이고  $\nu$ 값이 0인 경우는 단어중심의 채점 시스템이 되고 반대로  $\mu$ 값이 0이고  $\nu$ 값이 10인 것은 구문의미트리 중심의 채점 시스템이 된다. 나머지는  $\mu$ 값과  $\nu$ 값의 가중치 값에 따른 복합 채점 시스템이 된다.

##### - 분석 3

복합 채점 시스템과 단어 중심 채점 시스템, 구

문의미트리 중심 채점 시스템의 성능을 가중치 별로 그래프로 표현한 것이 [그림 7]이다.



[그림 8] 가중치 값에 따른 채점 성능

그래프에 나타난 것과 같이 가장 좌측의 시스템이 단어 중심의 채점 시스템이고 가장 우측이 구문의미트리 중심의 채점 시스템이다. 그리고 가운데 시스템들이  $\mu$ 값과  $\nu$ 값에 따른 복합 채점 시스템이다. 그림에서 알 수 있듯이 구문의미트리 중심의 채점 시스템이 단어 중심의 채점 시스템보다 나은 결과를 보였다. 이것은 주관식 문항에 대한 답변의 채점은 단어보다 구문의미트리가 더 중요한 개념과 의미를 가지고 있음을 증명하고 있다. 그리고  $\mu$ 값이 3이고  $\nu$ 값이 7인 복합 채점 시스템이 가장 좋은 성능을 보여주고 있음을 알 수 있었다.

#### 4. 결론 및 논의

본 연구는 주관식 문항 채점을 위해 구문의미 분석이 필요함을 제시하였고 구문의미분석의 결과인 구문의미트리를 비교하는 비교기를 이용하여 구문의미트리 중심의 채점 시스템을 제안하였다. 또한 단어 중심의 채점과 구문의미트리 중심의 채점의 복합 채점 시스템을 제안하였다. 본 연구의 실험 결과 구문의미트리 비교기를 이용한 구문의미트리 중심의 채점성능(0.983336)이 단어 중심의 채점성능(0.96869) 보다 나음을 알 수 있었다. 또한 채점성능 0.985495의 복합 채점 시스템도 구축할 수 있었다. 본 연구로 구문의미분석과 함께 구문의미트리 비교기를 이용한 구문의미트리 중심의 채점이 효과 있음을 알 수 있었다.

본 연구는 컴퓨터 전공분야의 지식에 대한 개념정의와 장단점 내용 설명을 요구하는 서술식



유형의 문항을 대상으로 하였다. 즉, 주제를 논하는 논술 유형이 아닌 기본적 지식을 논리적으로 설명하는 과학기술 분야의 주관식 채점에 적합한 것이다. 따라서 본 연구의 구문의미분석, 구문의미트리 비교기, 구문의미트리 비교기를 이용한 채점 방법은 과학기술 분야의 주관식 문항 채점에 바로 적용하여 채점의 질을 향상할 수 있을 것으로 기대된다. 또한 문서의 유사도를 계산하는 분야에도 활용할 수 있을 것이다. 앞으로 구문의미분석의 질을 향상할 필요가 있고 주관식 문항 유형별 처리 방법과 복문 또는 문장간의 연결관계 등을 고려한 채점 시스템의 연구가 필요하다.

## 참 고 문 헌

- [1] 정상목 · 한병래 · 송기상(2005). LSA를 이용한 서술형 주관식 평가 시스템의 설계 및 구현. **한국정보교육학회**, 9(2), 288-298.
- [2] 조정현 · 정현기 · 박찬영 · 김유섭(2009). BLEU를 활용한 단기 서술형 답안의 자동 채점. **한국HCI학회 학술대회**, 606-610.
- [3] 조우진 · 오정석 · 이재영 · 김유섭(2005). 의미 커널과 한글 워드넷에 기반한 지능형 채점 시스템. **정보처리학회논문지A**, 12-A(6), 539-546.
- [4] 강원석(2007). 구문-격의미 정보를 이용한 주관식 문항 채점 시스템 설계 및 구현. **한국컴퓨터교육학회 논문지**, 10(5), 61-69.
- [5] 강원석(2017). 주관식 문항 채점에서의 구문의미트리 비교 시스템에 대한 연구. **한국컴퓨터교육학회 논문지**, 20(6), 79-88.
- [6] 김동기 · 홍동권(2004). 주관식 평가를 위한 웹기반 온라인 평가 시스템의 구현 및 적용. **정보교육학회논문지**, 8(2), 1-9.
- [7] 이재영(2011). 대규모 주관식 문항에서 유사한 답 채점 정보관리 시스템. **한국정보기술학회논문지**, 9(2), 187-196.
- [8] 조지민 · 김경훈(2014). 서답형 문항 온라인 채점 시스템의 개발과 적용. **한국컴퓨터교육학회 논문지**, 17(2), 39-51.
- [9] 천민아 · 서형원 · 김재훈 · 노은희 · 성경희 · 임은영(2015). 준지도학습 방법을 이용한 한국어 서답형 문항 반자동 채점. **인지과학**, 26(2), 147-165.
- [10] Claudia Leacock & Martin Chodorow(2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities*, 37, 389-405.
- [11] Attali, Y. & Burstein, J.(2006). Automated Essay Scoring With e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3).
- [12] 최윤정 · 성태제(2006). 영어 논술 채점 컴퓨터 프로그램의 비교·분석. **교육평가연구**, 19(1), 145-160.
- [13] 박희정 · 강원석(2003). 유의어 사전을 이용한 주관식 문항 채점 시스템 설계 및 구현. **한국컴퓨터교육학회 논문지**, 6(3), 207-216.
- [14] 강원석(2011). 질의문 유형 분석을 통한 서답형 자동 채점 시스템. **한국콘텐츠학회논문지**, 11(2), 13-21.
- [15] 박일남 · 강승식 · 노은희 · 김명화 · 성태제(2013). 정답 템플릿 작성 방식에 의한 한국어 서답형 문항 자동채점 시스템. **정보과학회논문지 : 컴퓨팅의 실제 및 레터**, 19(12), 630-636.
- [16] 강승식 · 장은서(2016). 학생 답안 분석과 정답 템플릿 생성에 의한 한국어 서답형 문항의 자동채점 시스템. **정보과학회 컴퓨팅의 실제 논문지**, 22(5), 218-224.
- [17] 허정만 · 박소영(2013). 투표 기반 서술형 주관식 답안 자동 채점 모델의 설계 및 구현. **한국컴퓨터정보학회논문지**, 18(8), 17-25.
- [18] 천민아 · 김창현 · 김재훈 · 노은희 · 성경희 · 송미영(2016). 기계학습 분류기의 예측확률과 만장일치를 이용한 한국어 서답형 문항 자동 채점 시스템. **정보처리학회논문지/소프트웨어 테이터공학**, 5(11), 527-534.
- [19] 김재훈 · 선충녕 · 홍상욱 · 이성욱 · 서정연 · 조정미(1999). KTAG99: 새로운 환경에 쉽게 적용하는 한국어 품사 태깅 시스템. **제11회 한글 및 한국어정보처리 학술대회논문집**, 99-105.
- [20] 강원석 · 노주환 · 제환주 · 조흠 · 황세연 · 정부천(2007). 검색엔진을 한 키워드 관련어 추출기의 설계 및 구현. **한국컴퓨터교육학회**

2007년도 동계 학술대회 논문집, 241-246.

- [21] 강원석(2015). 구문의미트리 비교기를 이용한 유사문서 판별기. 한국콘텐츠학회논문지, 15(10), 636-646.



## 강원석

1985 경북대학교  
전자공학과(공학학사)  
1988 한국과학기술원  
전산학과(공학석사)

1995 한국과학기술원 전산학과(공학박사)  
1995~현재 안동대학교 컴퓨터교육과 교수  
관심분야: 컴퓨터교육, 자연어처리, 정보검색  
E-Mail: wskang@anu.ac.kr