

# 교육용 슬라이드의 표현적 난이도 측정에 관한 연구

김성찬\*, 송사광\*, 이문용\*\*

\*한국과학기술정보연구원 의사결정지원연구실

\*\*한국과학기술원 지식서비스공학대학원

e-mail : sckim@kisti.re.kr, esmallj@kisti.re.kr, munyi@kaist.ac.kr

## A Study on Representation Difficulty Assessment of Educational Presentation Materials

Seongchan Kim\*, Sa-Kwang Song\*, Mun Y. Yi\*\*

\*Decision Support Technology Lab., KISTI

\*\*Graduate School of Knowledge Service Engineering, KAIST

### 요 약

대학과 같은 고등교육 현장에서 슬라이드는 수업 또는 세미나에 많이 활용되는 매체이다. 최근에는 SlideShare 등의 슬라이드 전용 공유 플랫폼까지 등장하며 온라인상에 더 많은 교육용 슬라이드가 축적되고 있다. 이 연구에서는 이러한 슬라이드 형태의 교육 자료에 대해 인식하기 쉽고 어려운 정도인 표현적 난이도를 자동으로 측정하는 기법을 제안한다. 제안하는 60 개의 자질을 활용하여 기계학습 모델을 구축하고 표현적으로 고 난이도와 저 난이도의 슬라이드를 효과적으로 구분한다. 정밀하게 파악된 난이도 정보는 콘텐츠 선택에 있어 사용자 편의성을 획기적으로 증대시켜 줄 수 있다.

### 1. 서론

수 세기 동안 텍스트의 가독성(Readability; Difficulty of Reading)을 측정하기 위한 시도가 계속되어 왔다[1]. 가장 전통적인 방법은 문서 안의 어휘(Lexicon)들의 특성과 관련된 단순한 수식으로 문서의 난이도를 측정하는 방법이다[2][3]. 최근에 여러 자질(Feature)을 이용하여 기계학습 방법론을 활용한 텍스트 문서의 난이도를 측정하려는 연구가 시도되었다[4]. 한편 웹 페이지의 시각적 복잡도(Visual Complexity)를 측정하려는 연구가 수행돼 왔다. 웹 페이지의 시각적 복잡도는 사용자가 웹 페이지를 볼 때 얼마나 쉽게 인지할 수 있는지에 관한 정도이며, 최근의 연구들은 웹 페이지의 구조나 레이아웃, 색감 등의 자질을 고려한 기계학습 방법을 활용하여 이를 자동으로 측정하려 하였다[5][6].

하지만 슬라이드는 텍스트뿐만 아니라 그림, 수식, 표 같은 다양하고 복잡한 객체들이 포함된 청중을 대상으로 하는 발표자료로 주로 쓰이는 매체이다. 청중은 슬라이드마다 다른 구성과 편집으로 인해 내용을 받아들이는 데 있어 쉬움과 어려움을 다르게 느낄 수 있고 이는 슬라이드 문서의 난이도에 영향을 미친다. 이 연구에서는 이를 표현적 난이도(Representation Difficulty)라고 정의하고 이를 측정하는 새로운 방법론을 제안한다.

본 연구에서는 학습 슬라이드의 표현적 난이도 측정을 문제를 이진 분류(Binary Classification) 문제로 간주하고 고 난이도와 저 난이도의 슬라이드를 구분한다.

기존의 사용자 조사[7]에서 밝혀진 난이도 요소(Factor)를 기반으로 60 개의 난이도 관련 자질을 제안하고 슬라이드에서 자동으로 추출한다. 전문가집단으로부터 표현적 난이도 정보를 얻은 학습 데이터가 이용하여 분류기(Classifier)를 학습 시킨 후 테스트 데이터로 성능을 측정하였다.

본 논문의 의의는 다음과 같다.

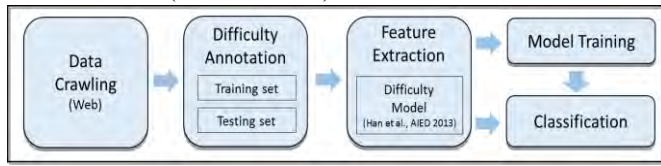
- 1) 교육용 슬라이드에 대해 표현적 난이도를 정의하고 이를 측정하는 방법론을 처음으로 제안하였다.
- 2) 표현적 난이도에 영향을 미치는 요소들에서 자동으로 추출 가능한 자질의 추출 방법을 고안하였다.
- 3) 표현적 난이도 평가 문제를 이진 분류 문제로 보고 제안하는 자질들을 이용하여 고 난이도와 저 난이도의 슬라이드를 효과적으로 구분하였다.

이 논문은 2 장에서 슬라이드의 표현적 난이도 측정 방법을 소개하고 3 장에서는 실험 과정과 결과를 제시한다. 그리고 4 장에서 결론을 내린다.

### 2. 교육용 슬라이드의 표현적 난이도 측정 기법

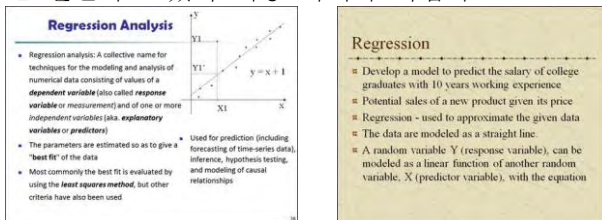
이 연구는 그림 1 에서 보는 바와 같이 5 단계의 절차를 거친다. 첫 번째 단계로 데이터 세트를 확보하기 위해 웹에서 교육용 슬라이드를 수집한다. 그 후 정답 세트(Ground Truth Set)를 구축하기 위해 전문가 집단으로부터 난이도 정보를 부여하는 주석

(Annotation)을 한다. 이후 기계학습을 위해 훈련 세트(Training Set)의 슬라이드에서 자질(Feature)을 추출하고 기계학습 모델을 학습시킨 후에 모델을 구축하고 테스트 세트(Test Set)에서 자질을 추출한 후에 난이도에 따라 분류(Classification)를 하였다.



(그림 1) 연구 절차

그림 2는 같은 주제(Regression Analysis)를 다루고 있는 교육용 슬라이드의 예를 보여주고 있다. 왼쪽의 슬라이드는 청중이나 독자가 내용을 받아들이기 쉽게 만들어진 즉 표현적으로 난이도가 낮은 슬라이드이며, 반대로 오른쪽의 슬라이드는 내용을 쉽게 받아들이기가 힘든 표현적 난이도가 높은 슬라이드이다. 왼쪽 슬라이드의 경우 내용 설명을 위해 관련된 수식과 그래프를 사용하고 중요한 부분에 굵은 텍스트를 사용하는 등 내용 이해를 도울 많은 노력을 기울이고 있으나 오른쪽 슬라이드의 경우 내용을 단순히 텍스트로만 전달하고 있어 내용 이해가 어렵다.



(그림 2) 표현적 난이도가 낮은 슬라이드(좌)와 높은 슬라이드(우)의 예

한 선행연구에서 슬라이드 난이도 관련하여 어떤 요소들이 난이도에 영향을 미치는지 알아보기 위하여 사용자 조사를 실시하였다[7]. 이 연구에 따르면 사용자들은 7개 분류의 49개 요소가 난이도에 영향을 미친다고 하였다. 본 연구에서는 이 선행연구에서 밝힌 난이도 요소들을 바탕으로 기계학습을 위한 자질을 고안 및 개발하였다. 표 1은 본 연구에서 사용한 5개 분류의 60개(측정하지 않은 2개 분류를 제외, 요소당 복수의 자질이 파생된 경우도 있음)의 교육용 슬라이드 난이도 측정에 관한 자질을 보여주고 있다.

<표 1> 난이도 측정을 위한 자질과 분류

분류	설명 및 자질(Feature)
Detailedness (10)	Factors that represent how comprehensible and concrete the slides are. (preHighlight, numHighlight, avgNumHigh, preExample, preFormula, preTable, preImg, preExtLink, preAnim, topicDen)

Structural Completeness (9)	Factors that represent how well-structured to be easily understood the slides are. (preSummary, preBullet, preNumbering, preTOC, preQ&A, numComp, compDistEasy, compDistHard, compDistRatio)
Readability (11)	Factors that indicate how well the text is comprehensible. (ARI, Flesh, Kincaid, SMOG, numStopWord, fracStop, avgTermLength, avgIDF, PPLM Easy, PPLMHard, PPRatio)
Length (16)	Factors that capture the size of the presentation. (numSlide, relNumSlide, numTerm, avgNumTerms, numImg, avgNumImg, numTable, avgNumTable, numFormula, avgNumFormula, numExample, avgNumExample, numExtLink, avgNumExtLink, numAnim, avgNumAnim)
Formatting Style (14)	Factors that capture the appearance of slides. (numFontSize, avgFontSize, maxFontSize, minFontSize, defFontColor, numFontColor, avgNumFontColor, numFontName, avgNumFontName, motherTongRatio, conFontFace, conFontSize, conIndLevel, conBackColorTemp)

### 3. 실험

데이터 세트를 확보하기 위하여 고등학교 과학과 교육과정(2015년 개정)에서 지구과학, 생명과학, 정보, 에너지와 환경 등의 교과목에 속한 교육 단원명 91개를 확보하여 크롤링의 키워드로 삼았다. 단원명과 한정자(Quantifier)를 조합하여 질의(예, “광합성” filetype:ppt site:ac.kr)를 만들고 구글<sup>1</sup>을 이용하여 대학교와 고등학교 사이트의 슬라이드만 크롤링하였다. 최종적으로 교육용 자료가 아닌 슬라이드(행정 문서)를 제외하고 고등학교 사이트에서 201개, 대학교 사이트에서 957개를 확보하였다.

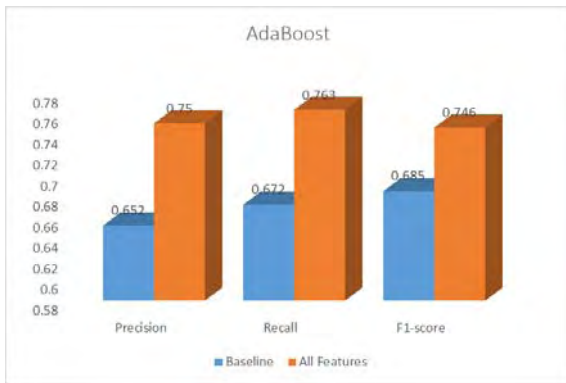
<표 2> 주석(Annotation) 후 최종 데이터 세트 (Ground Truth Set)

	저 난이도(Easy)	고 난이도(Hard)
샘플수	101	45

확보한 슬라이드에서 각각 100개씩 임의로 선택하여 표현적 난이도 정답세트를 구축하였다. 한 슬라이드당 5명의 전문가(해당과목을 이수한 대학교 3,4학년 및 대학원생)로 하여금 5점 척도(-2: 가장 어려움; 2: 가장 쉬움)로 표현적 난이도를 부여하도록 하였으며 5개의 수치를 평균하여 최종 난이도 점수를 산출하였다. 이때, 평균점수가 -0.2에서 0.2사이의 슬라이드는 정답세트에서 제외하였다. 최종적으로, 표 2에서 보는 바와 같이, 표현적으로 저 난이도(Easy)의 101개, 표현적 고 난이도(Hard)의 45개 슬라이드를 확보하였다. 평가자들의 일치도를 알아보기 위해 ICC(Interclass Correlation Coefficient)[8]를 측정하였으며 0.62(Good)를 기록하였다.

<sup>1</sup> http://google.com

제안하는 자질(표 1)을 추출하기 위해 POI Extractor<sup>2</sup>를 사용하여 슬라이드에서 텍스트, 그림 등 모든 객체를 추출한 후에 자질의 값을 계산하였다. 기계학습 모델로는 Weka Tool<sup>3</sup>의 AdaBoost(number of iteration: 10, weight threshold: 100, classifier: Random Forest)를 선택하였으며 기계학습 시 5-fold cross validation 을 실시하였다. 평가지표로 Precision, Recall, F1-score 를 사용하였으며 대조군(Baseline)으로는 Ou et al.의 웹 페이지의 시각적 복잡도 측정 방식[6]을 사용하였다. 그림 3 은 분류성능을 보여주고 있다. 제안한 자질들을 모두 고려했을 때 대조군 방법보다 F1-score 에서 8.9%의 성능 향상을 보였다. 구체적으로, 자질의 변별력을 판단하는 방법 중 하나인 정보 이득(Information Gain) 값이 높은 자질들인 단어의 평균 길이(avgTermLength-Readability)나 단어의 양(numTermLength), 사용한 글꼴의 평균 개수(avgNumFontName-Formatting Style) 등이 표현적 난이도를 구분하는데 주요한 역할을 하는 것으로 파악되었다. 표 4 에서는 각 집단 별 분류 성능을 보여주고 있다. 저 난이도 집단의 성능이 고 난이도 집단에 비해 높은 것을 알 수 있다. 이는 샘플 수의 차이에서 비롯된 것으로 보이며 샘플 수가 더 확보되면 더 좋은 성능을 보일 수 있을 것으로 기대된다.



(그림 3) 분류 성능

<표 4> 각 집단 별 분류성능

	AdaBoost		
	Precision	Recall	F1-score
저 난이도(Easy)	0.778	0.901	0.835
고 난이도(Hard)	0.655	0.422	0.514

4. 결론

본 연구에서는 슬라이드 형태의 교육자료의 표현적 난이도를 자동으로 분류하는 방법에 관하여 기술하였다. 본 연구는 사용자 조사에 기반을 두어 자질 개발에 개연성을 높였으며 제안한 자질이 효과적으로 난이도를 분류하는 데 기여함을 증명하였다. 앞으로 슬

라이드의 표현적 난이도뿐 아니라 내용적 난이도까지 고려된 종합적 난이도 또한 측정이 가능한 기법을 개발할 예정이다.

5. 사사

본 연구는 산업통상자원부 및 한국 산업 기술 평가 관리원의 산업핵심기술개발사업 (지식서비스) 일환으로 수행하였음. [10052955, 현장전문가의 경험지식 획득 및 활용을 위한 경험지식플랫폼 개발 연구]

참고문헌

- [1] K. Collins-Thompson, "Computational assessment of text readability: A survey of current and future research," *ITL - Int. J. Appl. Linguist.*, vol. 165, no. 2, pp. 97-135, 2014.
- [2] R. FLESCHE, "A new readability yardstick.," *J. Appl. Psychol.*, vol. 32, no. 3, pp. 221-233, Jun. 1948.
- [3] E. Dale and J. S. Chall, "A Formula for Predicting Readability," *Educ. Res. Bull.*, vol. 27, no. 1, pp. 11-28, 1948.
- [4] S. E. Petersen and M. Ostendorf, "A machine learning approach to reading level assessment," *Comput. Speech Lang.*, vol. 23, no. 1, pp. 89-106, Jan. 2009.
- [5] E. Michailidou, S. Harper, and S. Bechhofer, "Visual Complexity and Aesthetic Perception of Web Pages," in *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, 2008, pp. 215-224.
- [6] O. Wu, W. Hu, and L. Shi, "Measuring the Visual Complexities of Web Pages," *ACM Trans. Web*, vol. 7, no. 1, p. 1:1-1:34, 2013.
- [7] K. Han, M. Y. Yi, G. Gweon, and J.-G. Lee, "Understanding the Difficulty Factors for Learning Materials: A Qualitative Study," in *AIED*, 2013, pp. 615-618.
- [8] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychological Assessment*, Vol 6(4), pp. 284-290, 1994.

<sup>2</sup> <https://poi.apache.org/>

<sup>3</sup> <http://www.cs.waikato.ac.nz/ml/weka/>