Using Ontology-based Information Extraction for Subject-based Auto-grading

Olawande Daramola, Ibukun Afolabi, Ibidapo Akinyemi and Olufunke Oladipupo
Department of Computer and Information Sciences, Covenant University, Ota, Nigeria

Keywords: Ontology, Ontology-based Information Extraction, Automatic Essay Scoring, Subject-based Automatic Scoring.

Abstract: The procedure for the grading of students' essays in subject-based examinations is quite challenging particularly when dealing with large number of students. Hence, several automatic essay-grading systems have been designed to alleviate the demands of manual subject grading. However, relatively few of the existing systems are able to give informative feedbacks that are based on elaborate domain knowledge to students, particularly in subject-based automatic grading where domain knowledge is a major factor. In this work, we discuss the vision of subject-based automatic essay scoring system that leverages on semi-automatic creation of subject ontology, uses ontology-based information extraction approach to enable automatic essay scoring, and gives informative feedback to students.

1 INTRODUCTION

Student assessment task is usually challenging particularly when dealing with a large student population. The manual grading procedure is also very subjective because it depends largely on the experience and competence of the human grader. Hence, automated grading solutions have been provided to alleviate the drudgery of students' assessments. According to Shermis and Burstein (2013), notable systems for Automatic Essay Scoring (AES) include IntelliMetric, e-Rater, c-Rater, Lexile, AutoScore CTB Bookette, Page, and Intelligent Essay Assessor (IEA).

However, most of the existing AES systems have to be trained on several hundreds of scripts already scored by human graders, which are used as the gold standard from which the system learn the rubrics to use for their own automatic scoring. This procedure can be costly, and imprecise considering the inconsistent and subjective nature of human assessments. Also, most of the AES do not use elaborate domain knowledge for grading, rather they either use statistical or machine learning models or their hybrids, which limits their ability to give informative feedbacks to students on the type of response expected from based on the course content (Brent et al., 2010).

In this work, we present the vision of a subject-based automatic essay grading system that uses ontology-based information extraction for students' essay grading, and provides informative feedback to students based on domain knowledge. In addition, our approach attempts to improve on existing AES architectures for subject-based automatic grading by enabling the semi-automatic creation of relevant domain ontologies, which reduces the cost of obtaining crucial subject domain knowledge. Semi-automatic creation of domain ontology is particularly useful for subject grading where the only valid basis for assessment of students' responses is the extent of their conformity to the knowledge contained in the course content (Braun et al., 2006). Ontology as the deliberate semantic representation of concepts in domain and their relationships offers a good basis for providing more informative feedbacks in AES.

Hence, the intended contribution of our proposed approach stems from the introduction of ontology learning framework into AES as a precursor to providing informative feedbacks to students. Typically, our proposed approach employs ontology-based information extraction which uses basic natural language processing procedures – tokenization, word tagging, lexical parsing, anaphora resolution -, and semantic matching of texts to realise automatic subject grading.

The remaining parts of this paper are described

Daramola O., Afolabi I., Akinyemi I. and Oladipupo O..
Using Ontology-based Information Extraction for Subject-based Auto-grading.
DOI: 10.5220/0004625903730378
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2013), pages 373-378
ISBN: 978-989-8565-81-5

as follows. Section 2 contains a background and a review of the related work. In Section 3, we present the core idea of our approach, while Section 4 describe the process of ontology learning from domain text, and some of the essential aspects of our AES architecture. We conclude the paper in Section 5 with a brief note and our perspective of further work.

2 BACKGROUND AND RELATED WORK

Research on the viability of automatic essay scoring (AES) for student assessments have been undertaken since the 1960s, and several techniques have been used. The first AES, called *Project Essay Grade* (PEG) (Page, 1968) was implemented using multiple regression techniques. Some other methods that have been used for AES include: Latent Semantic Analysis (LSA) – *Intelligent Essay Assessor* (IEA) (Landauer and Laham, 2000); Natural Language Processing (NLP) - *Paperless School free-text Marking Engine* (PS-ME) (Mason and Grove-Stephenson, 2002), *IntelliMetric* (Elliot, 2003), e-Rater (Burstein, 1998); Machine Learning and NLP - LightSIDE, AutoScore, CTB Bookette (Shermis and Burstein, 2013); text categorization – (Larkey, 1998), CRASE, Lexile Writing analyzer (Shermis and Burstein, 2013), Bayesian Networks - *Bayesian essay testing system* (BETSY) (Rudner and Liang, 2002); Information Extraction (IE) - SAGrader (Brent et al., 2010); Ontology-Based Information Extraction (OBIE) - Gutierrez et al. (2012). Experimental evaluation of many of these AES also revealed that their scores have good correlation with that of human graders. However, majority of these systems cannot be used for short answer grading. An exception to this is IntelliMetric by (Elliot, 2003). A major drawback of many of these AES is that they have to be trained with scripts graded by human graders (usually in hundreds) for them to learn the rubrics to be used for text assessment. The human graded scripts serve as the gold standard for the evaluation, despite the fact that human judgments are known to be inconsistent and subjective. A more accurate basis for evaluation should be the fitness of student's response to the knowledge that must be expressed according to the course content. Also, they lack provision for subject-specific knowledge which limits their applicability to various subject domains, hence they are mostly for grading essays written in specific major languages. Therefore, they lack ability to provide informative feedbacks that stems from domain knowledge that can be useful to both students and teachers (Brent et al., 2010); (Chung and Baker, 2003).

In the category of short answer grading systems are examples such as c-rater (Leacock, C., and Chodorow, 2003), which is based on NLP; SELSA (Kanejiya et al., 2003) which is based on LSA and context-awareness; and Shaha and Abdulrahman, (2012) which is based on integrating Information Extraction (IE) technique and Decision Tree Learning (DTL).

The use of semantic technology for AES, which is the focus of our work, is relatively new, as very few approaches have been reported so far in the literature. The SAGrader (Brent et al., 2010) implements automated subject grading by combining pattern matching and use of semantic networks for domain knowledge representation. The system is able to provide limited feedback by identifying domain terms that are mentioned by students. SAGrader has limited expressiveness because a semantic network was used instead of an extensive ontology for domain knowledge representation. He et al. (2009), reported the use of latent semantic analysis, BLUE algorithm and ontology to provide intelligent assessment of students' summaries. Castellanos-Nieves et al., (2011) reports an automatic assessment of open questions in eLearning courses by using a course ontology and semantic matching. However, the ontology was manually created. Also, Gutierrez et al., (2012) used OBIE to provide more informative feedback during automatic student assessment by using an ontology that was manually created. However, creating ontology manually is a costly exercise, which is not realistic for large subject domains that will require large and complex ontologies. Also, creation of ontology requires high technical expertise which is not common.

Hence, our approach intends to improve on existing OBIE approaches by enabling the semi-automatic creation of the ontology from domain text, and giving informative feedbacks that stem from domain knowledge to students, and even teachers for both short answers grading and long essays. The form of feedback will entail misspellings, correct and incorrect statements, and incomplete statements, and structure deficiency in sentence constructions.

3 OVERVIEW OF THE APPROACH

The core idea of the proposed approach is outlined as sequence of offline and run-time activities as follows.

(i) Select relevant subject domain text and information sources that can be used to train a lexical tagger, such as OpenNLP or Stanford NLP tree tagger – this will enable greater accuracy of natural processing activities such as part-of-speech tagging of words that will subsequently be encountered in students' scripts and teacher's marking guide.

(ii) Create an ontology for the subject domain semi-automatically from textual information sources of the domain such as text books/book chapters or lecture notes. In cases, where such relevant domain ontology already exists, select it to use for the grading process by importing it into the proposed architectural framework.

(iii) Create a meta-model schema of the marking guide of the subject prepared by the examiner, which will be used by the auto-grading system as basis to associate questions to corresponding responses by students. The meta-model is typically a graph-based data structure (see Fig 1.) that describes the arrangement of the questions in the exam/test that is used as a logical template to map a student's response to corresponding sections of the marking guide on a question-by-question basis. It captures the description of each question (q1-q4) in terms of the number of its sub-parts (a, b, c ...), its unique identification (id), type of response expected (R) – classified into 3 categories, *list*, *short essay,* and *long essay* -, and the mark allocated (M) to the question. The description of a question in the meta-model primarily determines the type of semantic treatment that is applied when extracting information from a student's response.

(iv) Collate students' responses to specific questions and pre-process the student's response by conducting spelling checks, identifying wrong punctuations, and noting right or wrong use of domain concepts. Keep track of all corrected instances, which will be included in the feedback to students.

(v) Extract information from student's response based on pre-defined extraction rules depending on the type of expected response as contained in the marking guide meta-model. Evaluate the lexical structure of each sentence in the response to a question by performing subject-predicate-object (SPO) analysis of each sentence in order to extract the subject (noun), predicate (verb), and object (noun). For a correct statement, the extracted subject, and object should correspond to specific concepts in the domain ontology, either, in their exact form, root form or synonym forms, while the predicate should be valid for the concepts in the sentence based on the taxonomy, and axioms of the underlying ontology.

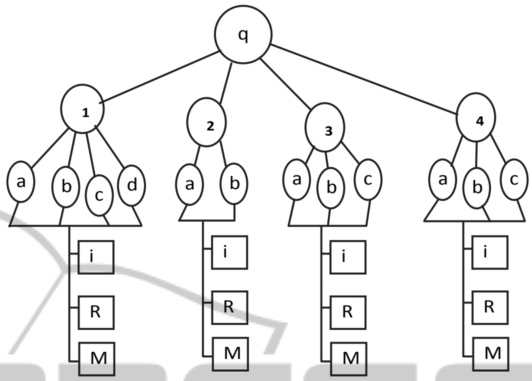

Figure 1: A schematic view of the marking guide meta-model.

(vi) Perform text semantic similarity matching of the information extracted from student's response, and the content of the marking guide. Two possibilities exist, depending on the expected response to a question. First, for questions where short, or long essay response are expected, extract rules using the *<concept>* *<predicate>* *<concept>* pattern to analyse each sentence of the answer to that question as contained in the marking guide. The extracted rules are then matched semantically with the result of SPO analysis of student's response to determine similarity and then scoring. Second, for questions where the type response expected is a list, extract a bag of concepts from the marking guide and compare with the bag of words from student's response using a vector space model to determine semantic similarity.

(vii) Execute an auto-scoring model based on the degree of semantic similarity between a student's response to a question (C_s) and the teacher's marking guide (C_m) using the domain ontology for reasoning. The semantic similarity $sim(C_s,C_m)$ in the interval [0-1] will be the basis for assigning scores – e.g. $sim(C_s,C_m) > 0.7$ = full marks; $0.5 \geq sim(C_s,C_m) \leq 0.7$ = 75% of full marks; $sim(C_s,C_m) < 0.5 = 0$.

(viii) Accumulate score obtained per question and repeat steps (iv) - (viii) until all questions have been graded.

4 ONTOLOGY LEARNING FROM SUBJECT DOMAIN TEXT

Our approach for ontology learning emulates the ruled-based procedure for extracting seed ontology from raw text as employed by (Omoronyia et al., 2010; Kof, 2004). The steps of the ontology building process are described as follows:

***Document Preprocessing*:** This is a manual procedure to ensure that the document from which ontology is to be extracted is fit for sentence-based analysis. The activities will include replacing information in diagrams with their textual equivalent, removing symbols that may be difficult to interpret, and special text formatting. The quality of pre-processing of a document will determine the quality of domain ontology that will be extracted from such source document.

***Automatic Bracket Trailing*:** This is a procedure to identify sentences/words that are enclosed in bracket within text and to treat them contextually. Usually in the English language, brackets are used in text to indicate reference pointers e.g. ("Fig 2") or ("see Section 4") or to embed supplementary text within other text. The bracket trailing procedure ensures that reference pointers enclosed in brackets are overlooked and that relevant nouns that are enclosed in brackets are rightly associated with head subject or object that they refer to depending on whether the bracket is used within the noun phrase (NP) or verb phrase (VP) part of the main sentence. Consider the sentence: *"E-Commerce (see Fig. 1) involves the exchange of goods and services on the Internet based on established electronic business models (such as Business-to-Business, Business-to-Customer, and Customer-to-Business)"*. Bracket trailing will ensure that the reference "see Fig. 1" is overlooked, while the noun subjects "*Business-to-Business", "Business-to-Customer", "Customer-to-Business"* are related to object *electronic business models.* Relations derived via bracket trailing are semantically related to relevant subject/object in text by using a set of alternative stereotypes such as <refers to>, <instance of> or <same as> depending on the adjective variant used with an extracted noun. The domain expert that is creating the ontology is prompted to indicate his preference.

***Resolution of Term Ambiguity*:** This involves a semi-automated process of discovering and correcting ambiguous terms in textual documents using observed patterns in a sentence parse tree (Omoronyia, et al., 2010). To do this, the observed pattern in a particular sentence parse tree is compared with the set of collocations (words frequently used together) in the document in order to identify inconsistencies. When the usage of a word in a specific context suggests inconsistency, then the relevant collocation is used to substitute it, in order to produce an ontology that is more representative of the subject domain.

***Subject Predicate Object (SPO) Extraction*:** This procedure uses a natural language parser to generate a parse tree of each sentence in the document in order to extract subjects, objects and predicates. The structure of each sentence clause consists of the Noun Phrase (NP) and Verb Phrase (VP). The noun or variant noun forms (singular, plural or proper noun) in the NP part of a sentence is extracted as the subject, while the one in the VP part is extracted as the object. The predicate is the verb that relates the subject and object together in a sentence.

***Association Mining*:** This explores the relationship between concepts in instances where a preposition other than a verb predicates relates a subject and an object together. A prepositional phrase consists of a preposition and an object (noun). Automatic association mining is a procedure that detects the existence of a prepositional phrase and relates it with the preceding sibling NP. Example *"E-Commerce as a form of online activity is gaining more prominence"*. Here, *E-Commerce* is the subject, while "as *a form of online activity*" is a prepositional phrase containing the object "*online activity*". Association mining will recognize the inferred relationship between "*E-commerce*" and "*online activity*" and associate them together by using the generic stereotype <relates to>.

***Concept Clustering*:** This entails eliminating duplications of concepts, and relationships in all parsed sentences. Also, concepts are organized into hierarchical relationships based on similarity established between concepts.

The semi-automatic procedure for ontology enables the domain expert to revise the seed ontology through an ontology management GUI interface in order to realize a more usable, and more expressive ontology. From the ontology management GUI, the domain expert can create ontological axioms – restrictions such as allValuesFrom, somevaluesFrom, hasValues, minimum cardinality and maximum cardinality in order to facilitate inference of new interesting knowledge.

5 KEY COMPONENTS OF THE AES FRAMEWORK

The architecture of our proposed AES will be composed of an integration of components and procedures that will help to realize automatic grading via a sequential workflow. It accommodates a series of activities that can be classified as offline and online activities. The offline activities include: training the natural language POS tagger on domain text to aid recognition of domain specific terms, the architecture will afford an interface to import the domain text, and train the POS tagger. An ontology learning and management module that leverages algorithms for shallow parsing and middleware algorithms implemented by Stanford NLP[1], and Protégé OWL 2[2] will used to perform ontology extraction from domain text in order to create domain ontology for the subject domain concerned semi-automatically.

The other essential components of the architecture are described as follows:

Meta-model Engine: it automatically transforms a teacher's marking guide into an intermediate formal representation that forms the basis for semantic comparison with a student's response to questions. It has an interface where the teacher will input metadata information for specific questions – its unique id, type of response expected, and mark allocated -, and the answer to each questions. Based on these information, the marking guide meta-model will be automatically created.

*Information Extractor***:** This component will implement a Semantic Text Analyser (STA). The STA will serve as the semantic engine of the AES system. It will employ a combination of natural language processing procedures, and domain knowledge to make sense out of a student's response based on some pre-defined extraction rules. STA will perform semantic text analysis such as tokenization of text, term extraction, word sense disambiguation, and entity extraction using the domain ontology and WordNet.

Auto-Scoring Engine: This component will perform semantic matching of the contents that have been extracted from the students' response to specific questions and the equivalent marking guide meta-model representation of specific questions. It will use a pre-defined scoring model (see Section 3) to determine score allocated to the a student's response to a question

Resources Repository (RR): This refers to the set of data, knowledge, and open source middleware artefacts that will enable the semantic processing capabilities of the AES framework. All other components of the AES framework leverage on the components of the RR to realise their functional objectives. A brief overview of the role of elements of the RR is given as follows:

Domain Ontology: the domain lexicon that encapsulates knowledge of the subject to be graded. It is used to enable the extraction of Information from students' responses.

WordNet – An English language lexicon used for semantic analysis.

MySQL – A database management system used to implement data storage in the AES framework. MySQL's capability for effective indexing, storage, and organisation will aid the performance of the AES in terms of information retrieval, and general usability.

Protégé OWL API – A Java-based semantic web middleware that is used to facilitate ontology query and management, and ontology learning from text.

Pellet – An ontology reasoner that support descriptive logics reasoning on domain ontology components.

Standford NLP – A Java-based framework for natural language processing. It will provide the set of APIs that will be used by the Information extractor component of the AES framework.

6 CONCLUSIONS

In this paper, we have presented the notion of ontology-based information extraction framework for subject-based automatic grading. Relative to existing approaches, the benefit of the proposed framework is the semi-automatic creation of domain ontologies from text, which is capable of reducing cost of subject automatic essay grading significantly. In addition, our proposed framework will improve on existing efforts by enabling informative feedbacks to students. It also affords greater adaptability, because it allows for grading in several subject domains, once there is suitable domain ontology, and a relevant marking guide. However, the proposed approach relies primarily on the existence of a good quality ontology, which means the domain expert may still need to do some enhancements after the semi-automatic creation

[1] http://www-nlp.stanford.edu/software/lex-parser.shtml
[2] http://protege.stanford.edu/download/registered.html

process in order to realise a perfect ontology. Nevertheless, the additional effort required will definitely be less than the cost of creating a good ontology from the scratch. As an ongoing work, we intend to realize the vision of the framework is the shortest possible time, and to conduct some evaluations using a University context.

REFERENCES

Braun, H., Bejar, I., and Williamson, D., 2006. Rule based methods for automated scoring: Application in a licensing context. In D. Williamson, R. Mislevy and I. Bejar (Eds.), *Automated scoring of complex tasks in computer based testing*. Mahwah, NJ: Erlbaum.

Brent E., Atkisson, C. and Green, N., 2010. Time-Shifted Online Collaboration: Creating Teachable Moments Through Automated Grading. In Juan, Daradournis and Caballe, S. (Eds.). *Monitoring and Assessment in Online Collaborative Environments: Emergent Computational Technologies for E-learning Support,* pp. 55–73. IGI Global.

Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., Harris, M., 1998. Automated Scoring Using A Hybrid Feature Identification Technique. *In Proc. COLING '98 Proceedings of the 17th international conference on Computational linguistics*, pp. 206-210.

Castellanos-Nieves D., Fernandez-Breis J.T., Valencia-Garcia R., Martinez-Bejar R., Iniesta-Moreno M., 2011. Semantic Web Technologies for supporting learning assessment. *Information Sciences,* 181 (2011) pp. 1517–1537.

Chung, G., and Baker, E., 2003. Issues in the reliability and validity of automated scoring of constructed responses. In M. Shermis and J. Burstein (Eds.), *Automated essay grading: A cross-disciplinary approach*. Mahwah, NJ: Erlbaum.

Elliot, S., 2003. A true score study of 11^{th} grade student writing responses using IntelliMetric Version 9.0 (RB-786). Newtown, PA: Vantage Learning.

Gutierrez, F., Dou, D. and Fickas, S., 2012. Providing Grades and Feedback for Student Summaries by Ontology-based Information Extraction. *In. Proceedings of the 21st ACM international conference on Information and knowledge management* (CIKM '12), pp. 1722-1726.

He, Y., Hui, S., and Quan, T., 2009. Automatic summary assessment for intelligent tutoring systems. *Computers and Education*, 53(3) pp. 890–899.

Kanejiya D., Kumar, A., and Prasad, S., 2003. Automatic evaluation of students' answers using syntactically enhanced LSA. *Proceedings of the HLTNAACL 03 workshop on Building educational applications using natural language processing*-Vol. 2, pp. 53–60.

Kof, L., 2005. An Application of Natural Language Processing to Domain Modelling - Two Case Studies. *International Journal on Computer Systems Science Engineering*, 20, pp. 37–52.

Landauer, T.K., and Laham, D., 2000. The Intelligent Essay Assessor. *IEEE Intelligent Systems*, 15(5), pp. 27–31.

Larkey, S., 1998. Automatic essay grading using text categorization techniques. *Proceedings of the 21st annual international ACM conference on Research and development in information retrieval (SIGIR '98)*. pp. 90-95.

Leacock, C., and Chodorow, M., 2003. C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities*, Vol. 37 (4), pp. 389-405.

Mason, O. and Grove-Stephenson, I., 2002. Automated free text marking with paperless school. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.

Omoronyia, I., Sindre, G., Stålhane, T., Biffl, S., Moser, T., Sunindyo, W.D., 2010. A Domain Ontology Building Process for Guiding Requirements Elicitation. In Requirements Engineering: Foundation for Software Quality, *Lecture Notes in Computer Science* Volume 6182, pp. 188-202.

Page E.B., 1968. The Use of the Computer in Analyzing Student Essays. *Int'l Rev. Education*, Vol. 14, pp. 210–225.

Rudner, L. M. and Liang, T., 2002. Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment*, 1 (2), 3–21.

Shaha T. A. and Abdulrahman A. M., 2012. Hybrid Approach for Automatic Short Answer Marking. *Proceeding of Southwest Decision Sciences Forty Third Annual Conference*, New Orleans, pp. 581 -589.

Shermis M., J. Burstein (2013). *Handbook of Automated Essay Evaluation, Current Applications and New Directions*, Routledge, New York.