Software That Grades Handwritten Essays May Boost Comprehension, Too

January 14 2008

Computer scientists in the University at Buffalo's School of Engineering and Applied Sciences have been working with their colleagues in UB's Graduate School of Education to develop a computational tool that not only dramatically reduces the time it takes to grade children's handwritten essays, but that also may help boost students' reading comprehension skills.

The software has special relevance to the school systems and teachers involved in administering the standardized English Language Arts exams that are given every year, usually in January, by public school systems in every state. This month, every New York school district will administer these assessments to their students in grades three to eight.

The National Science Foundation recently awarded the UB researchers a $100,000 grant to develop new algorithms that could eventually allow computers to take over the grading of children's handwritten essays.

The UB team's preliminary results with the software are scheduled for publication in the February/March issue of Artificial Intelligence. The paper was published earlier in the online version of the journal.

"It surprised us that we were able to do as well as we did, especially since this was our first attempt," said Sargur N. Srihari, Ph.D., SUNY Distinguished Professor in the UB Department of Computer Science and Engineering and principal investigator on the project.

The project focused on handwritten essays obtained from eighth graders in the Buffalo Public Schools who responded to this question from a New York State English Language Arts exam: "How was Martha Washington's role as First Lady different from that of Eleanor Roosevelt?"

Three hundred of the essays were scored by human examiners and used as a "gold standard" against which 96 computer-scored essays were judged.

Essays were graded on a scale of 0-6, with six being the highest score.

In 70 percent of cases, the UB researchers reported, the computer program graded the essays within one point of those assigned by human examiners.

The UB research tackles two significant artificial intelligence problems, said Srihari, director of UB's Center of Excellence in Document Analysis and Recognition (CEDAR), the world's largest research center devoted to developing new technologies that can recognize and read handwriting.

"We wanted to see whether automated handwriting recognition capabilities can be used to read children's handwriting, which is essentially uncharted territory," he said. "Then we took it one step further to see if we could get computers to score these essays like human examiners."

In the pilot study, the essays were first scanned into a computer. Each line of text was broken down into individual words. In this step, the system's goal was word recognition, which it accomplished using contextual information from the rest of the sample, the answer rubric and the question.

Once the majority of words were recognized, the essay was turned into a digital text file.

For the automated scoring step, the UB researchers used an artificial neural network approach.

"In this method, the system 'learns' from a set of answers that were scored already by humans, associating different values or scores with different features in the essays," explained Srihari.

Computational tools designed to evaluate essays that are typed, not handwritten, already exist, Srihari explained.

"But these are all based on electronic text that the test-taker types in, using a computer keyboard," he said. "In this case, we are working toward developing a computational tool to read and evaluate the many thousands of handwritten essays written by schoolchildren as part of statewide mandated reading comprehension tests."

The sheer speed with which the program works -- literally seconds per essay -- is the most obvious advantage, the UB researchers said.

Handwritten essays are an important part of every standardized reading comprehension test given in every state. But because grading all of those handwritten essays is such a huge task requiring many hours of work by human examiners, students who take the exam in January do not find out how they did until almost the end of the spring semester.

"Judging this quantity of handwritten essays is very laborious," said Srihari. "It would be nice to automate this process so perhaps students could take the test in May, having received more instruction, and then have the results in June."

And while some teachers may be wary of computers' ability to properly grade essays, James L. Collins, Ed.D., professor in the UB Department of Learning and Instruction and a co-investigator, is quite confident.

While he noted that human examiners might still be necessary for grading on very specific criteria, the majority of evaluations could probably be done just as well by computers.

"Computational linguistics has made great leaps over the past decade and it turns out that for judging the overall quality of a paper, computers are indeed as reliable as human graders," Collins said.

That's an important development, he said, because writing practice and feedback from readers are the key aspects of learning to write at every grade level.

"The problem is, 'How do teachers respond helpfully to all of the writing produced by their students?'" he said. "Right now, teachers spend a lot of time getting their students ready for these standardized tests, then the students take the exam and get their scores back months later. With computer scoring, students could get back their scores much faster at a time when the results can still be addressed. The assessment scores wouldn't just be going into a 'black hole.'"

The software program developed at UB was 'trained' to evaluate essays based on six specific writing traits: ideas, organization, word choice, sentence structure, voice and conventions like spelling, usage and punctuation.

Collins said that the software now under development could be used as an important teaching tool.

"We envision a program where a student would handwrite an essay, scan

it into the computer, which would then 'read' it and analyze it for the specific traits we trained it to evaluate," he said.

That feedback would be available immediately to both teacher and student as a typed essay, which has been analyzed for the six traits, allowing for more fruitful lessons on how to edit and revise, Collins said.

The software program also provides new opportunities for education researchers like Collins, who is working with colleagues at UB on a three-year, $1.5 million project called Writing Intensive Reading Comprehension funded by the Institute of Education Sciences at the U.S. Department of Education. The study involves more than 2,000 fourth and fifth graders in 10 low-performing urban schools. So far, Collins said, the results show that students can improve their reading abilities significantly through the use of assisted writing.

"Once a handwritten essay has been 'read' by a computer, we can ask the computer to look for certain features of the writing so that we can spot general patterns and discover what kids are having trouble with," Collins continued.

Co-authors on the Artificial Intelligence paper with Srihari and Collins are Janina Brutt-Griffler, Ed.D., associate professor in the UB Department of Learning and Instruction; Rohini Srihari, Ph.D., professor of computer science and engineering at UB; Harish Srinivasan, a doctoral candidate at CEDAR, and Shravya Shetty, a former graduate student at CEDAR, now employed by Google.

Source: University at Buffalo

Citation: Software That Grades Handwritten Essays May Boost Comprehension, Too (2008,