Automatic classification of incorrect answers to differentiation questions using Potential Response Tree

Yasuyuki Nakamura[1] Saburo Higuchi [2] Kentaro Yoshitomi[3] Yoshinori Miyazaki[4] Yuko Ichikawa[5] Takahiro Nakahara[6]

Abstract: In this study, the authors focus on the Potential Response Tree (PRT), with which users may define rules to classify answers and give feedbacks/partial credit by how they are incorrect or incomplete. Various wrong answers need to be considered for each question, depending on the combination of elementary functions that make up the question. We categorize differentiation questions and propose an algorithm to classify incorrect answers automatically using a PRT according to the type of questions.

Keywords: Potenaial Response Tree; Differentiation question; Classification of incorrect answers

1 Introduction

Mathematics e-Learning system is one of the online assessment systems, in other words, computer-aided assessment (CAA) systems with which students can provide mathematical expressions as answers by calculation, and the responses can be automatically assessed correct or incorrect. A question type of the mathematics e-Learning system is entirely different from traditional question types of CAA; for example, true or false, single-answer or multiple-choice question type, because students do not answer by choosing the answer from multiple potential options but provide solutions determined by calculation. STACK[Sa13, ST] , with which we investigate incorrect answers in this study, is an example of a mathematics e-Learning system. There are some other examples such as Möbius Assessment[Mo] (formerly Maple T.A.), WeBWorK[We], Numbas[Nu], MATH ON WEB[Ka16] and so on.

Because answer data are not restricted to one of the potential options in the mathematics e-Learning system, there could be various types of answers depending on the students' levels of understanding. Let us consider a question of differentiation $\frac{d}{dx}(2x+1)^3$as an example. The correct answer is $6(2x+1)^2$, but an answer $3(2x+1)^2$ is likely to be provided because

[1] Nagoya University, Graduate School of Informatics, Nagoya, Japan nakamura@nagoya-u.jp
[2] Ryukoku University, Faculty of Advanced Science and Technology, Otsu, Japan hig@math.ryukoku.ac.jp
[3] Osaka Prefecture University, Faculty of Liberal Arts and Sciences, Sakai, Japan yositomi@las.osakafu-u.ac.jp
[4] Shizuoka University, College of Informatics, Hamamatsu, Japan yoshi@inf.shizuoka.ac.jp
[5] National Institute of Technology, Tokyo Ccollege, General Education Department, Hachioji, Japan yuko@tokyo-ct.ac.jp
[6] Sangensha, LLC., Chitose, Japan nakahara@3strings.co.jp

doi:10.5281/zenodo.4915995

of the lack of knowledge about the chain rule of the differentiation of composite functions. The answer could also be $24x^2 + 24x + 6$ to the same question. In this case, it is guessed that the student might have expanded the given expression at first and differentiated it next. The answer is algebraically equivalent to the correct answer, but it is hard to know if the student knows differentiating composite functions. Therefore, it is important not only to assess students' answers as correct or incorrect but to know what kind of answers students generated. Provided incorrect answers can give hints to show what knowledge students lack and further can be categorized by classification. The provision of thus obtained clues would be quite helpful to novice teachers on the site. The result of our preliminary study[Na19], which we will briefly review, has motivated us to promote this study.

2 Brief review of incorrect answer analysis

Let us briefly review our previous study[Na19]. The online assessment was conducted in a calculus course at one of the colleges of the National Institute of Technology, Japan, in 2017. In the class, differentiation and integration of functions of one variable, including trigonometric, exponent, and logarithmic functions, were taught. There were 103 students who took the course, and they addressed online tests as assignments at home. All data are stored in the STACK database, and we collected data of questions of differentiation of composite functions. All data are labeled depending on the status of answers submitted by students.

As a first attempt, we investigated a question of differentiation of $\sin^n x$ or $\cos^n x$ where n is selected randomly from 2, 3, and 4 in a test. There were 74 attempts at the test by 42 students, and 100 answers were collected for analysis. Note that four attempts were omitted because they were not completed. The test itself was basic and most of the students submitted correct answers. However, there were various types of incorrect answers, and we found that they could be classified into "typical" types. Our first attempt was to find typical mistakes students tend to make by classification of answers. There were many types of answers, and Table 1 shows a summary of the classification of students' answers for the question of differentiation of $\sin^n x$ or $\cos^n x$. In the case of numbers 1, 2, 3, and 4 on the table, students may understand the rule of differentiation of trigonometric function. Still, they did not even care the question was on the differentiation of a composite function. Those who submitted answers numbers 5 and 6 understand the basics of differentiation of trigonometric functions and polynomial functions, but they do not understand the differentiation of composite functions. Those types of answers were the most typical. The student re-submitted the correct answer on his or her the second attempt.

Therefore, it is important to investigate students' problem-solving process, and we can know students' understanding process. In order to realize it, it is essential to detect what kind of incorrect answers students submit to the question, and automatic classification of wrong answers should be a fundamental procedure.

No	Math expressions for differentiation	Student's answer	# of answers
1	$\sin^n x$	$\cos^n x$	3
2	$\cos^n x$	$(-\sin x)^n$	3
3	$\sin^n x$	$n\cos^n x$	2
4	$\cos^n x$	$(-\sin x)^n$	3
5	$\sin^n x$	$n\cos^{n-1} x$	4
6	$\cos^n x$	$(-\sin x)^{n-1}$	2

Tab. 1: Typical incorrect answers to the differentiation question $\frac{\mathrm{d}}{\mathrm{d}x}\sin^n x$ or $\frac{\mathrm{d}}{\mathrm{d}x}\cos^n x$

3 Automatic classification of incorrect answers using Potential Response Tree

In order to classify incorrect answers to differentiation questions, we decided to categorize questions depending on the form of mathematical expressions of questions, functions used in questions, and the relation (or operation) of functions. A series of online testing was conducted in calculus class. The 50 differentiation questions were picked up, and we categorized them into specific types. The categorization of questions asked in the class is summarized in Table 2. Mathematical expressions are examples because parameters such as coefficients and indices are randomly selected. Incorrect answers are submitted depending on the specific type of questions. We classify incorrect answers by using Potential Response Tree (PRT), a powerful feature of STACK, to assess students' answers. Our goal is to build the most versatile PRT possible to classify incorrect answers to as many types of questions as possible.

As a first step, we picked up the question of differentiation of product of polynomial functions that is expressed by the form $f_1 \cdot f_2$ with $f_1 = a_1x^{m_1} + a_2$ and $f_2 = a_3x^{m_3} + a_4$ in which a_1, a_2, a_3, a_4, m_1 and m_3 are randomely determined. Then we try to build PRT for the questions and regrade students' answer by applying the PRT.

3.1 Question variables

We define "Question variables" when editing a STACK question as follows. Coefficients and indices are randomly selected and $f_1(x)$ and $f_2(x)$ are defined as a function of x using `lambda` function. Then mathematical expression that should be differentiated is defined by the product of the two functions, and the differentiation of it is defined as `dpx` that will be referenced as a correct answer.

```
a1:rand(5)+2;
m1:rand(2)+2;
a2:(rand(3)+3)*(-1)^rand(2);
a3:rand(3)+1;
```

Category	Maxima (like) expression	Form	f@g (@; operator)				function					
			fg	1/g	f/g	f(g)	x^a	sin	cos	tan	a^x	ln
polynomial1	3x^6+3x^5-4x^2-2x	f1					4					
polynomial2	(4x^2+4)(3x^4-4)	f1*f2	1				4					
polynomial3	(5x^4+3x+3)^(40)	f1(f2)				1	4					
polynomial4	(5-4x)^3(x+3)^3	f1(f2)*f3(f4)	1			2	6					
power1	x^(1/2)	f1					1					
power2	(3-4x)^(-3/5)	f1(f2)				1	3					
power3	(6x+7)/(3x-2)	f1/f2			1		4					
power4	3 (x^3+4)^(-4)	f1(f2)				1	3					
power5	(4x+4)/(x^(1/2))	f1/f2			1		3					
power6	(-4x^2-6x-2)^(-3)	f1(f2)				1	4					
power7	x^(1/2)(5x+4)	f1*f2	1				3					
trigonometric1	cos(2x+4)	f1(f2)				1	2	1				
trigonometric2	(sin(x))^3	f1(f2)				1	1	1				
trigonometric3	(tan(x))^2	f1(f2)				1	1			1		
trigonometric4	1/(tan(x)+2)	1/f1		1			1			1		
trigonometric5	(cos(4x+5))^4	f1(f2(f3))				2	3		1			
trigonometric6	(sin(x)+1)^(1/2)	f1(f2)				1	2	1				
trigonometric7	x^3cos(5x-6)	f1*f2(f3)	1			1	3		1			
trigonometric8	(e^(2x)tan(x)+1)^4	f1(f2(f3)*f4,f5)	1			2	3			1	1	
trigonometric9	x^(6x)	NA	-	-	-	-	-	-	-	-	-	-
trigonometric10	(x-10)^4(x-5)^3/((x+4)^3)	f1(f2)*f3(f4)/f5(f6)	1		1	3	9					
exponential1	e^(6x+5)	f1(f2)				1	2				1	
exponential2	4^(-3x-5)ln(x)	f1(f2)*f3	1			1	2				1	1
exponential3	(x^4-15x)e^x	f1*f2	1				2				1	
exponential4	(e^(3x)-6)^(-1/4)	f1(f2(f3), f4)				2	3				1	
exponential5	(e^(6x)-5)^(1/2)	f1(f2(f3), f4)				2	3				1	
exponential6	x^5e^(4x+5)	f1*f2(f3)	1			1	3				1	
exponential7	e^(3x)/(e^(3x)+5)	f1(f2)/(f3(f4),f5)			1	2	3				2	
exponential8	e^(3x+4)/(x^5)	f1(f2)/f3			1	1	3				1	
exponential9	3^x+4^(-4x-5)	f1,f2(f3)				1	2				2	
exponential10	4^(3x+4)ln(x)	f1(f2)*f3	1			1	2				1	1
logarithmic2	ln(4x-5)	f1(f2)				1	2					1
logarithmic4	x^(1/2)ln(2x-5)	f1*f2(f3)	1			1	3					1
logarithmic5	ln(4x+5)/x	f1(f2)/f3			1	1	3					1
logarithmic6	1/(ln(x)+6)	1/f1		1			1					1
logarithmic7	(ln(x)+2)^4	f1(f2)				1	2					1
logarithmic8	ln((x+3)^2/x)	f1(f2(f3)/f4)			1	2	4					1
logarithmic9	log_2|2x+4|	f1(f2)				1	2					1
logarithmic10	log_10(x^4)	f1(f2)				1	1					1

Tab. 2: Categorization of differentiation questions asked in the class

```
m3:rand(2)+1+m1;
a4:(rand(3)+3)*(-1)^rand(2);
f1:lambda([x],a1*x^m1+a2);
f2:lambda([x],a3*x^m3+a4);
px:f1(x)*f2(x)
dpx:diff(px,x);
```

3.2 Potential response tree

We design a PRT assuming the following four types of incorrect answer candidates: $f_1'(x) \cdot f_2(x)$, $f_1(x) \cdot f_2'(x)$ and $f_1'(x) \cdot f_2'(x)$. In order to build PRT, the following are defined as “Feedback variables”.

```
df1:lambda([x],diff(f1(x),x));
df2:lambda([x],diff(f2(x),x));
```

```
isZero:is(ans1=zero);
isCorrect:is(expand(df1(x)*f2(x)+f1(x)*df2(x)-ans1)=0);
isdff:is(expand(df1(x)*f2(x)-ans1)=0);
isfdf:is(expand(f1(x)*df2(x)-ans1)=0);
isdfdf:is(expand(df1(x)*df2(x)-ans1)=0);
```

`df1` and `df2` are defined as $f_1'(x)$ and $f_2'(x)$ that are differentiation of $f_1(x)$ and $f_2(x)$. We define a boolean variable `isCorrect` that is true when students' answer is equall to correct answer $f_1'(x)f_2(x) + f_1(x)f_2'(x)$ and false when students' answer is incorrect. The Answer test `AlgEquiv` is usually used to determine whether students' answer is correct or incorrect, but it sometimes fails to evaluate equivalence check of students' answer and correct answer, especially when the function is complicated. We are planning to apply the same PRT to other types of questions such as trigonometric function, exponent function, and so on in order to generalize PRT. Therefore we expand the function `df1(x)*f2(x)+f1(x)*df2(x)-ans1` and evaluate it as zero or not. One of the node of PRT using the `isCorrect` function is shown in Figure 1. "Answer note" is recorded as "isCorrect" when the node is true and "not_isCorrect". As we see later, the answer note is used in "STACK response analysis". Note that "Node 1" moves to the next "Node 2" regardless of the true or false result of "Node 1". In a similar way, nodes from "Node 2" to "Node 5" are defined using the functions `isZero`, `isdff`, `isfdf` and `isdfdf`.

As an example, when a student answer is $a_1m_1a_3m_3x^{m_1+m_3-2}$ which is a incorrect answer of $f_1'(x)f_2'(x)$ type to the differentiation question $\frac{\mathrm{d}}{\mathrm{d}x}(a_1x^{m_1} + a_1)(a_3x^{m_3} + a_4)$, the answer notes "`not_isCorrect`, `not_isZero`, `not_isdf1f2`, `not_isf1df2` and `isdf1df2` are recorded. The procedure of detecting the type of incorrect answer by using PRT is schematically shown in Figure 2. The pink background color means the result of each answer test. For example, for the incorrect answer $f_1'(x)f_2'(x)$, only the result of answer test `isdfdf` is true and the answer note `isdf1df2` is returned. One of the features of our PRT is that it is a chain structure instead of a tree structure. Because of the structure, a new node can be easily added if it is necessary.

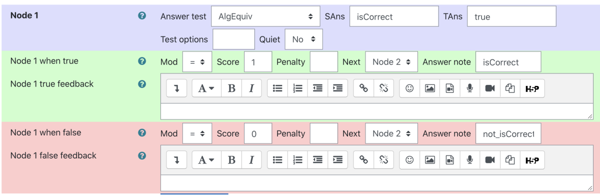

Fig. 1: One of the node of PRT that uses the `isCorrect` function.

4 Results of STACK response analysis

For the question $\frac{\mathrm{d}}{\mathrm{d}x}(a_1x^{m_1} + a_1)(a_3x^{m_3} + a_4) = \frac{\mathrm{d}}{\mathrm{d}x}f_1(x)f_2(x)$, the classification of answers was carried out and summarized in Table 3. 60 out of 107 correctly answered, and

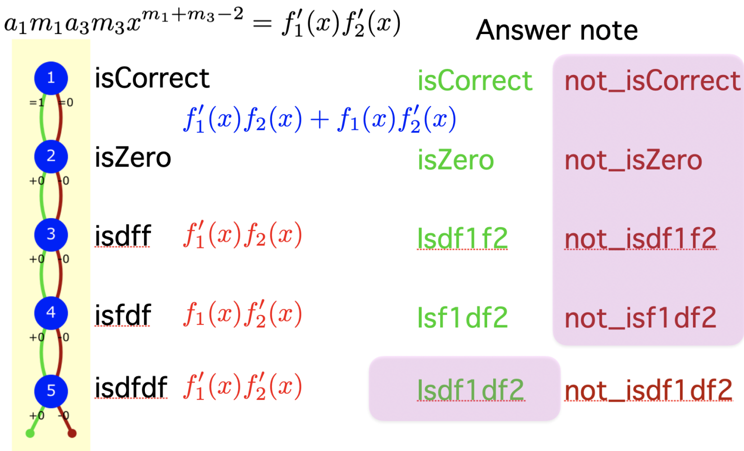

Fig. 2: Schematic figure that shows the assessment of student's answer $a_1m_1a_3m_3x^{m_1+m_3-2}$ to the differentiation question $\frac{\mathrm{d}}{\mathrm{d}x}(a_1x^{m_1} + a_1)(a_3x^{m_3} + a_4)$.

we attempted to classify the 47 incorrect answers into some types. Only 11 incorrect answer of type $f_1'(x)f_2'(x)$ were found. Other 36 incorrect answers can be considered to be miscalculated answers. According to the records of answer notes, only the function `isCorrect` is true for correct answers and only the function `isdf1df2` is true for incorrect answers of the type $f_1'(x)f_2'(x)$. On the other hand, none of the functions is true for unclassified answers.

As we expected, the incorrect answer of the type $f_1'(x)f_2'(x)$ is a typical wrong answer, and the number was just over 20% of the total of incorrect answers. However, nearly 70% of incorrect answers were not classified yet.

Answer note of PRT	Classification	#
isCorrect | `not_isZero` | `not_isdf1f2` | `not_isf1df2` | `not_isdf1df2`	Correct answer	60
`not_isCorrect` | `not_isZero` | `not_isdf1f2` | `not_isf1df2` | **isdf1df2**	$f_1'(x)f_2'(x)$	11
`not_isCorrect` | `not_isZero` | `not_isdf1f2` | `not_isf1df2` | `not_isdf1df2`	Other incorrect answer	36

Tab. 3: Results of STACK response analysis

5 Improving a PRT

In order to make more detailed classification possible, we define another function by using Maxima. Even if a student's answer does not match one of the three types of incorrect answers described above, when those kinds of terms are included in students answers, in order to detect those answers, we make new functions; `checkTerm`, `hasdff`, `hasfdf` and `hasdfdf`.

```
checkTerm(sterm,tterm): is(1=simplify(sterm/tterm));
hasdff: isdff or (not atom(ans1) and op(ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTerm(sterm,(df1(x)*f2(x)))),args(ans1))));
hasfdf:  isfdf or (not atom(ans1) and (ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTerm(sterm,(f1(x)*df2(x)))),args(ans1))));
hasdfdf: isdfdf or (not atom(ans1) and op(ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTerm(sterm,(df1(x)*df2(x)))),args(ans1))));
```

Furthermore, in order to realize that only coefficients difference can be detected, the function `checkTermC`, `isCdff`, `isCfdf`, `isCdfdf`, `hasCdff`, `hasCfdf` and `hasCdfdf` are defined.

```
isCdff: checkTermC(ans1,df1(x)*f2(x));
isCfdf: checkTermC(ans1,f1(x)*df2(x));
isCdfdf: checkTermC(ans1,df1(x)*df2(x));

hasCdff: isCdff or (not atom(ans1) and op(ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTermC(sterm,(df1(x)*f2(x)))),args(ans1))));
hasCfdf: isCfdf or (not atom(ans1) and op(ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTermC(sterm,(f1(x)*df2(x)))),args(ans1))));
hasCdfdf: isCdfdf or (not atom(ans1) and op(ans1)=op(x+y) and
apply("or",map(lambda([sterm],checkTermC(sterm,(df1(x)*df2(x)))),args(ans1))));
```

As a result, out of 36 unclassified incorrect answers, new 11 incorrect answers were detected. Our PRT characteristic is that classification can be carried out with a combination of multiple nodes that become true. For example, an incorrect answer $4x(2x^4+4)+2x^3(2x^2-3)$ to the question $\frac{\mathrm{d}}{\mathrm{d}x}(2x^2-3)(2x^4+4)$ is true for multiple nodes, in other words, two functions `hasdff` and `hasCfdf` are true.

6 Conclusion

We build a potential response tree (PRT) to classify incorrect answers into specific types by making full use of Maxima. As a first step, we build a PRT for the question of differentiation of product of polynomial functions. According to the "Answer note" of PRT, the classification was made using "STACK response analysis". Our PRT is not like a tree structure but a chain structure, and it is easily improved.

In this study, we only build PRT for the specific type of differentiation question. Still, we will make as general as possible PRT for many question types to classify incorrect answers. We already operate a question database system Mathbank [Na14], and we are planning to add model questions using general PRT to classify incorrect answers. Our goal is to know the understanding process of students on the basics of question-solving process study.

7 Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 18H01069.

Bibliography

[Ka16] Kawazoe, M.; Yoshitomi, K.: E-learning/e-assessment systems based on webMathematica for university mathematics education. MSOR Connections, 15(2), 2016.

[Mo] Mobius Assessment - Online Assessment System for STEM Courses | DigitalEd. https://www.digitaled.com/products/assessment/. accessed: 18/06/2021.

[Na14] Nakamura, Y.; Taniguchi, T.; Nakahara T.: Item Bank System for the Mathematics e-Learning System STACK. Electronic Journal of Mathematics & Technology, 8(5), 2014.

[Na19] Nakamura, Y.; Ichikawa, Y.; Miyazaki Y.; Yoshitomi K.; Higuchi S.; Nakahara T.: An Attempt to Analyze Mathematical Question-solving Processes Using STACK. In: Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), pp. 2210–2213, 2019.

[Nu] Numbas, Really versatile maths e-assessment. https://www.numbas.org.uk/. accessed: 18/06/2021.

[Sa13] Sangwin, C.: Computer aided assessment of mathematics. OUP Oxford, 2013.

[ST] STACK | The University of Edinburgh. https://www.ed.ac.uk/maths/stack/. accessed: 18/06/2021.

[We] Welcome to WeBWorK. http://webwork.maa.org/. accessed: 18/06/2021.