A data-driven approach for the identification of misconceptions in step-based tutoring systems

Joice Cazanoski Gomes[1], Patricia A. Jaques[1]

PPGCA - UNISINOS
São Leopoldo – RS – Brazil

cgjoice@gmail.com, pjaques@unisinos.br

Abstract. *Math errors are an important part of the learning process. For this reason, diagnosing them can help teachers and intelligent learning environments to choose the most appropriate type of assistance for the learner. In particular, the identification of learner misconceptions can be of special importance because they represent a misunderstanding of math concepts. In this context, this paper proposes the use of clustering algorithms to automatically identify algebra misconceptions from learners' algebra problem-solving steps in an intelligent learning environment. The computing platform is an intelligent tutoring system that assists students when solving linear equations step by step, by giving minimal and error feedback. The results showed that the model was able to identify some misconceptions already known in the literature, which illustrates the appropriateness of our approach. The automatic identification of misconceptions can help in the identification of new conceptual misunderstanding from a large dataset of math problem-solving, besides give valuable information for teachers and intelligent learning environments to adapt their instruction and assistance.*

1. Introduction

A misconception is defined as an error caused by a misunderstanding of a concept, and it has singular characteristics: the error repetition and the independence from numeric values. In Math teaching, misconceptions are inevitable, and they can be interpreted as an alternative way to find the correct answer. In this case, it is possible to analyze how students solve a math task step by step, trying to understand how the student assimilates theory and practice [MEC 1997].

Intelligent Tutoring systems that assists students when solving problems can be called **answer-based**, when they only give feedback for the final answer, or **step-based**, when the assistance is provided for each solving step. Studies based on meta-analyses [VanLehn 2011] have shown that step-based tutors are more effective than answer-based tutors, and the impact on learner's performance, compared to human one-to-one tutoring, is similar. One explanation for the better performance of the step-based tutors, in relation to the answer-based ones, is that they can infer a more detailed model of student's skills, including misconceptions, since they follow the entire process of solving a problem [VanLehn 2011].

The identification of misconceptions is essential to help the ITS to provide individual assistance and become more effective to students' learning. Although the automatic detection of misconceptions is still considered a hard task, it is directly connected to the

DOI: 10.5753/cbie.sbie.2020.1122

capacity of the artificial tutor to improve the students' knowledge in a specific domain [Elmadani et al. 2012]. Besides, this information can also be shared with teachers in real-time, allowing them to adapt their instruction.

In the research literature, several works have proposed methods to evaluate learners' answer patterns that are related to algebra misconceptions. However, these models or methods cannot cope with the variability of errors because either they use bug libraries, (e.g., [Russell et al. 2009]), limiting the identified errors, or they use the raw data without any modification, making it harder to identify patterns, like the related works cited in Section 2.

In this context, this work aims to automatically identify the occurrence of a misconception using problem-solving log data from a step-based algebraic intelligent tutor system. The data used in this work is from the ITS PAT2Math and is composed of 1,070 registers from wrongs steps submitted by students. Each register contains the student id, the equation difficulty level, the current step, the previous step, and the minimal feedback of the step (right or wrong). We are using educational data mining techniques, more explicitly, the clustering method, to identify equation patterns that represent misconceptions.

2. Related Works

[Russell et al. 2009] present a study that aims to analyze misconceptions in algebra, intending to recognize the effect of identifying misconceptions and the use of these data by teachers on learners performance. The authors used an existent online test tool (DAAS - Diagnostic, Algebra Assessment System) that aims to diagnose three known literature misconceptions: concept of variable, equality and graphing. Students also completed a pre- and a post-test containing similar questions related to the same misconceptions. Two metrics were defined to measure the students' performance on the tests: performance as the total number of correct answers, and misconceptions as the total number of items where the student selected the answer with misconceptions. The results showed evidence that the use of DAAS by teachers positively affects the identification of misconceptions and the learners' performance in algebra. However, the authors also presented some limitations: the number of students who presented misconceptions was too small concerning the sample, about 27%, decreasing the ability to identify effective differences between the groups. Besides, the study focused on just three mainly known misconceptions, limiting indirectly the number of possible students that could participate in the study.

[Feldman et al. 2018] searches on how to solve the problem of identifying unknown misconceptions in the literature. Thus, the work proposes a new form of identification, exploring the student's answer, and analyzing all possible paths to obtain such an answer. In the first experiment, the objective was to measure how many known misconceptions were identified. The obtained result was 70%, 28 of the 40 listed misconceptions were detected correctly. For the second experiment, the input was students' data collected by MetaMetrics. The data is composed of the answers of 296 students for 32 addition and subtraction problems. Solution groups with three or more solutions of the same student were defined, resulting in 868 groups, and 111 of these groups had at least one misconception. The results showed that the algorithm was able to rebuild the solution of 86% of the groups, being 77 of these classified as Accurate (the algorithm rebuilt the group misconception correctly) and Partially Accurate (the reconstruction did not represent the real solution process). The authors highlighted that the fact of the used learning system (MetaMetrics) only allow the entrance of the final solution to the problem was a limitation, since steps in the students' problem-solving that could help in the identification were omitted. The other limitation is that the proposed solution assumes that all of the same group's problems are solved the same way. In many cases, that does not happen, and random errors that are not caused by misconceptions generate algorithm noise.

The work by [Andersson and Johansson 2015] aims to develop an ITS for teaching math for children of 6 to 10 years old, and that can also identify misconceptions. They propose using clustering techniques for the errors to be grouped and to perform the misconception identification related to these errors. A first solution was developed using the algorithm k-means. However, the necessity of identifying the number of existing clusters in the data was the main problem faced in this first approach. The second implementation used the algorithm DBSCAN that solved the identified problem with the previous algorithm, but it still needed to inform some parameters. The authors performed one more change, using a fuzzy variant of DBSCAN (FN-DBSCAN), so the errors could have association degrees in different clusters, finding similarities among the error groups. The test with real users did not identify misconceptions, possibly because of the small number of data from different students. For a more accurate evaluation of the solution, the authors wanted to test the system with known misconceptions. For this reason, they created simulated users with known misconceptions. After twenty rounds, the system was able to identify 107 errors and one possible related misconception. After forty rounds, it identified 206 errors and three possible misconceptions. The author concluded that the ITS behaves better by analyzing users data in the long term.

[Elmadani et al. 2012] presents the use of data mining for misconception identification in the EER-Tutor ITS. Although it offers feedback to punctual mistakes. However, it cannot identify when those errors result from misconceptions. The authors utilized the FP-Growth algorithm to find possible misconceptions from the most frequent relations between incorrect answers in multiple-choice questions. The data is a list of 1135 answers logs. An expert analyzed the data to identify potential misconceptions related to the collected errors. The next step was to generate association rules with 0.9 minimum trust levels. The data processing performed by the algorithm generated 912 sets. On average, a restriction appeared in 44 sets. The more significant number of occurrences was 238, appearing in at least 6700 students' attempts, which was considered misconceptions. The authors began creating a results-based hierarchy, indicating that a violated constraint at the top of a hierarchy could mean that the rest of the branch is a misconception. Experts analyzed the results and considered it as satisfactory. As improvements, the authors suggest the identification that different restrictions can be similar in structure and have a relation with the same misconception.

2.1. Comparative analysis

The majority of related works were applied in the math domain, a well-defined domain. Besides, most recent works used machine learning algorithms for the identification of misconceptions, which demonstrates the viability of this approach for the problem of misconception detection.

Regarding the type and the source of the data used for misconception identification, [Russell et al. 2009] used a diagnosing test and [Andersson and Johansson 2015] simulated the data. [Feldman et al. 2018] used only the final answer of the math problems, which limits the results, because many misconceptions are in the solving steps. Also, [Feldman et al. 2018]'s work assumes that the equations found in the same group are solved in the same way, resulting in the wrong outputs. Our proposal is able to detect unknown misconceptions, is based on real data collected from students in the school setting and analyzes all steps of the problem-solving.

Besides, related work used raw data without any modification for misconception identification. Based on [Elmadani et al. 2012], which highlights that the misconceptions have a similar structure, we propose to improve the detection by an initial data processing. We normalize the equation data, changing numeric values to variables, becoming easier to identify patterns.

Works with automatic identification are considered those that do not need human intervention to understand and evaluate the results. Although automatic partial detection has superior technology support than non-automatic detection, it still needs expert intervention. Almost all related works need, in any point, some human intervention to identify the misconceptions, except [Andersson and Johansson 2015], who do not need human intervention, but their proposal has satisfactory results with simulated entries. As far as we are aware, our work is the first proposal of complete automatic detection with real user data.

3. Material and Methods

This work proposes a model to automatically identify misconceptions from students steps in step-based tutors for first-degree equations. The main original contribution of the proposed work is an automatic identification of the misconceptions, that is, the proposed model does not need human intervention to identify whether misconceptions are involved in a given solution and which is this misconception. Besides, the misconceptions are identified from all the steps of the problem-solving, and not only the final answer of a problem, which allows a more accurate identification.

One of the misconceptions characteristics is the repetition of the same error. Accordingly, the best way to identify misconceptions is to find groups of the same mistake; then, clustering methods are the best approach to solve this problem. In this work, we use the k-modes clustering algorithm, as our data is strictly categorical. The input data for developing and test the solution are from the ITS PAT2Math, a step-based ITS for teaching first-degree equations.

Our approach consists in normalizing the equation terms to have a homogeneous structure, changing numeric terms for constants. This way, the variation from equation elements do not affect the main structure, making it easier to cluster the data and the misconception identification.

3.1. Data

For the proposed work, we used log data from the ITS PAT2Math. The data consists of steps of first degree equation solving and the minimal feedback of the tutor (step was right or wrong). We have only retrieved registers in which the student entered a wrong answer.

For each step of the student, the ITS PAT2math stores a register in the database. Each register of a step contains the student id, the equation difficulty level, the current step, the previous step, and the minimal feedback of the step (right or wrong). To identify a misconception, we need to identify and evaluate the mistake made by a student. Consequently, we only need the wrong steps and its previous step. As the purpose of this work is to identify misconceptions, regardless of the student or the complexity, we started removing the other useless data from the dataset.

The first data extraction used to the first tests contained only 255 entries. After some tests, we got a second data extraction, which included 1070 entries, to use in our final model. We do not need to filter or classify the data because we use clustering. However, we decided to limit the number of terms we will use in each equation. For this, we explored the dataset, and we found six entries that were very big and unique, which means this data would probably result just in noisy in the results, and we removed these entries. The final dataset was composed of 1064 entries, from easy (Equation 1) to complex (Equation 2) levels of difficulty.

$$x + 4 = 9 \tag{1}$$

$$\frac{-5x + 255}{16} - 125x = -1000 + \frac{555x - 111x + 16}{8} \tag{2}$$

3.2. Pre-processing

Related work has used the input data without modifications (raw data) to identify misconceptions (see Section 2). However, in Mathematics, the structure of an equation or function is very important to identify these misconceptions. Besides, the repetition of an error is the main characteristic of a misconception and it is easier to find this repetition if we represent the equation in a more high-level representation.

Based on the mentioned facts, to have a better result, we decided to normalize the equations, changing the numeric values for variables. Table 1 contains the captions used to modify the values. We decided to keep the number cardinality because some well-known misconceptions in math literature depend on the cardinality of the number (e.g., a one-digit number divided to a two-digit number).

Table 1. Data transformation

Description	Example	Changed
Zero	0, 0x	z, zx
One digit number	1, 9, 4	a, b, c
Two digit number	10, 25, 99	aa, bb, cc
Negative number	-6, -22, -101	-a, -bb, -ccc
Unknown with multiplier	2x, 10x, -3x	ax, bbx, -cx

The Equation 3 represents an equation before the modification.

$$-2x - 10 + 4 = 15 - 5x \tag{3}$$

The Equation 4 represents the same equation after modification.

$$-ax + bb + c = aa - bx \tag{4}$$

The next step on data processing was to create a structure, separating the terms and operations of the equations in different parameters. Next, we defined a maximum number of terms for each entry; here, we identified these terms and operations and created the main structure. Close to each operator (for example, multiplication, sum, etc), there are two terms because we have some cases that have parenthesis, and we are treating them as terms. When an equation does not have all terms, we complete with **none**. Figure 1 presents the final arrangement for the previous and wrong steps.

Previous Step																
Left Term 1	Left Term 2	Left Operator 1	Left Term 3	Left Term 4	Left Operator 2	Left Term 5	Left Term 6	Left Operator 3	Left Term 7	Left Term 8	Left Operator 4	Left Term 9	Left Term 10	Left Operator 5	Left Term 11	Left Term 12
equal																
Right Term 1	Right Term 2	Right Operator 1	Right Term 3	Right Term 4	Right Operator 2	Right Term 5	Right Term 6	Right Operator 3	Right Term 7	Right Term 8	Right Operator 4	Right Term 9	Right Term 10	Right Operator 5	Right Term 11	Right Term 12

Wrong Step															
Left Term 1	Left Term 2	Left Operator 1	Left Term 3	Left Term 4	Left Operator 2	Left Term 5	Left Term 6	Left Operator 3	Left Term 7	Left Term 8	Left Operator 4	Left Term 9	Left Operator 5	Left Term 11	Left Term 12
equal															
Right Term 1	Right Term 2	Right Operator 1	Right Term 3	Right Term 4	Right Operator 2	Right Term 5	Right Term 6	Right Operator 3	Right Term 7	Right Term 8					

Figure 1. Previous and wrong step structure

3.3. The proposed Model

In this section, we present the steps followed in the creation of the proposed model, including selecting the algorithm, the implementation, and the parameters selection.

3.3.1. Algorithm selection

We decided to use the clustering method due to the repetition characteristic of misconceptions. Besides, as clustering is an unsupervised approach, it allows to detect unknown misconceptions. After the pre-processing, our data became exclusively categorical, limiting our choices.

Focusing on the algorithms used by the related works, we decided to follow the same approach used for [Andersson and Johansson 2015]. However, they used the k-means algorithm, and for the proposed work, we used the categorical variant of k-means, called k-modes. Besides, differently from them, who used simulated data, we used real data of students who have practiced equation solving in PAT2Math.

3.3.2. Parameters selection

The k-modes requires some parameters, and an adequate selection of these parameters can have a strong impact on the results. We adopted the following steps to choose the parameters:

- **init**: we ran each parameter, except the fixed ones, in a small dataset to choose the best initialization method. We manually evaluated the results, choosing the one that generates less noise. The best choice was the method proposed by [Huang 1998], which selects the most frequent categories as centroids;
- **n_init**: this parameter sets the number of times the algorithm will run with different centroid seeds, choosing the best run, in terms of cost, as output. We set this parameter as default (10);
- **n_clusters**: for choosing the optimal number of clusters, we used the elbow method. We ran the algorithm in our final dataset using a range from two to fourteen clusters. The final result for the optimal number of clusters (k) is 8 clusters, based on the cost and diversity of our clusters.

3.4. Output format

Using k-modes, we clustered each register defined on the dataset and after we formatted the results, resulting in our misconceptions. Our final output is a pair of equations (the previous step and current wrong step), representing the misconception. We need both previous and current step, to be able to identify which misconception is involved in the solution and the cause. For example, if we only have the equation, flagged as wrong $x = 10 * 2$, we cannot identify what is wrong. However, if we also have the previous step, $x * 2 = 10$, it is possible to infer that the error is related to the incorrect application of the inverse operation for multiplication. Each cluster have a group of equations, and we get the two results with more occurrence, which means each cluster results in two misconceptions structures. With this output, we can already identify the misconceptions encountered in the dataset.

4. Results and Discussion

In this section we present the results obtained and an analysis based on the literature references and the set objectives for this work.

4.1. Results

For each running in the test phase, we observed the data distribution across the clusters. However, this distribution changed a little for each trial.

Then, we stored our output (see Section 3.4) for some executions; when we have an already added entry, we ignored it. Another behavior noticed in the test phase was that some results do not have a consistent structure — for example, Equation 5, which has a closing parenthesis without the opening. We eliminate these inconsistent results as they represent typos and not misconceptions.

$$x + bb - ccx) = aaa + bbb \tag{5}$$

Each cluster resulted, on average, two structures that we consider as a misconception. However, some results are not misconceptions. Figure 2 shows detailed results, showing the raw result and our analysis for each outcome.

Result		Evaluation
Previous Step	Wrong Step	
$x - b = aa$	$x = -aa$	Calculation Error
	$x = aa$	
$x + b = -aa$	$x = -a$	
$x + b = aa$	$x = a$	
$x = aa - b$	$x = aa$	
$(aax + bb) - (cc - ddx) = (-aa + bbx) - (-ccx + dd)$	$aax + bb - cc + ddx = -aa + bbx + ccx + dd$	Sign Rules misconception
$a * (x - c) - \frac{dx}{e} = a * \frac{x - c}{d}$	$ax - b - \frac{cx}{d} = \frac{ax}{b}$	Polynomial Multiplication misconception
$-x - bb = \frac{aa}{b}$	$-\frac{x}{b} = aa + bb$	Inverse Operations misconception
$x + bb = -a$	$x = -a + bb$	
$ax + bb = ax + bb$	$ax + bx = aa - bb$	
$x + b = aa$	$x + b - c = aa - b$	Unknown Isolation misconception
$-x - bbb = -aaa$	$-x - bbb + ccc = -aaa + bbb$	
$ax - bb = aa$	$ax + bb = aa + bb$	
$aax + bb = aax - bb$	$aax - bbx = -aa - bb$	No misconception or error
$(aax + bb) - (cc - ddx) = (-aa + bbx) - (-ccx + dd)$	$x = \frac{a}{b}$	Noisy
	$aax = -aa$	
$\frac{x}{bb} - \frac{ccc}{ddd} = \frac{-x + bb}{cc}$	$\frac{x}{bb} = -x$	

Figure 2. Results

4.2. Discussion

The proposed model was able to automatically, with no-human intervention, identify misconceptions, which was a limitation in most related works. Besides, our approach is able to identify unknown misconceptions and it is based on data from real students. Another differential of our approach, counterpointing [Andersson and Johansson 2015], is that it is able to identify misconceptions in a dataset from different users using clustering.

[Elmadani et al. 2012] does not apply a homogeneous structure in their proposal but was the first work to propose as future work to use this approach to identify misconceptions without human intervention. We applied the proposed approach in this work, and could observe that the results were improved in comparison to [Elmadani et al. 2012] and [Andersson and Johansson 2015] results, even without human intervention.

However, some aspects could be improved to make this model more accurate. As shown in Figure 2, we had some outputs with an inconsistent structure, mainly caused by parenthesis. This is a side effect of our decision to treat them as regular terms. So, as future work, we can analyze how to manage them as special terms to preserve the original structure. A possible solution for this is to use expression trees to represent the steps (equations).

Also, some results represent just a calculation error, not being classified as a misconception; [Feldman et al. 2018] reported the same limitation on his work.To solve this issue and as a future improvement to automatically generate a more precise report, we can create a library with our already known misconceptions, and also the ones which only represent an error, formatted with our structure together with a description, similar to our evaluation. Then, after each run, the program matches the results, returning a report with the found misconceptions. But, different from bug libraries, we will not discard the entries that do not match because they can represent a new unknown misconception.

We also identified some noisy results, but those need a more in-depth analysis to understand the root cause. This problem can be another side effect of the parenthesis terms and also about fractions, because we represented fractions as a division operation together with two terms in parenthesis, like *(a)/(b)*; in these cases, the parenthesis was not separated from the number. Another hypothesis is that some noise results from incomplete sentences in the dataset. But we will need to go deeper into the dataset to have more exact conclusions on it.

Finally, in this work, we applied the proposed model in a past dataset extracted from the ITS PAT2Math. As future work, we want to analyze the possibility of integrating it as a new feature within the ITS PAT2Math to provide even more individual learning. Also, it can also give a new way for the teachers to evaluate the students. For example, after a test, the ITS identifies a misconception for a student; it will be possible to personalize the tips and feedback for the student, trying to "fix" this misconception. For the teacher, it could have the possibility to access the data about the class or specific students and improve the lessons to have a better understanding. This way, the feature will help not just the students but also the teachers.

5. Conclusion

Mathematics teaching in Brazil has been below the level expected in recent years, and in many cases, this metric results from misconceptions since the beginning of learning *Sistema de Avaliação da Educação Básica* [SAEB 2017]. However, identifying these misconceptions is a hard task even to humans. Computer tutors are evolving to be able to identify them automatically. Nevertheless, in most cases, they still need a human intervention to have an accurate result. So, this work aimed to identify math misconceptions automatically and also be able to identify unknown misconceptions. Also, we used real data, consequently representing real misconceptions.

One of the main characteristics of a misconception is the repetition; for this reason, clustering is a feasible approach for these cases. Also, a misconception is not dependent on the numeric values in an equation. Then we proposed to normalize the values, changing the numeric values to variables. We applied the categorical clustering algorithm k-modes, and our results have shown that we can automatically identify these misconceptions. We have identified five misconceptions in our dataset, and other type of errors, such as calculation errors, which are not classified as a misconception.

However, we identified some flaws in our proposed structure, mainly caused by the way we treat parenthesis. We decided not to have a particular term to the parenthesis, which caused some inconsistencies in the results. Also, we have some unexpected results; like structures representing a calculation error, this can cause a misunderstanding about the results.

Consequently, as future works, we propose to review our structure and treat special terms, like parenthesis, by using expression trees. Also, we can add a “match system” to eliminate results that are calculation errors. Another possible future work is to integrate this model as a feature on the ITS PAT2Math, providing more individual teaching.

Agradecimentos

O presente trabalho foi realizado com apoio da FAPERGS (Processo 17/2551-0001203-8) e CNPq (processo 309218/2017-9).

References

Andersson, J. and Johansson, H. (2015). Using clustering in a cognitive tutor to identify mathematical misconceptions. *LU-CS-EX 2015-45*.

Elmadani, M., Mathews, M., and Mitrovic, A. (2012). Data-driven misconception discovery in constraint-based intelligent tutoring systems. In *Proceedings of the 20th International Conference on Computers in Education*, Singapore.

Feldman, M. Q., Cho, J. Y., Ong, M., Gulwani, S., Popović, Z., and Andersen, E. (2018). Automatic diagnosis of students' misconceptions in k-8 mathematics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 264, Montréal, Canada. ACM, Author.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 2(3):283–304.

MEC (1997). *Parâmetros Curriculares Nacionais 5ª a 8ª Séries*. MEC.

Russell, M., O'Dwyer, L. M., and Miranda, H. (2009). Diagnosing students' misconceptions in algebra: Results from an experimental pilot study. *Behavior research methods*, 41(2):414–424.

SAEB (2017). *Resultados Avaliação do Ensino Básico*. INEP.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221.