Trakya Üniversitesi Sosyal Bilimler Dergisi 487
Aralık 2024 Cilt 26 Sayı 2 (487-504)
DOI: 10.26468/trakyasobed.1435490 Araştırma Makalesi/ Research Article

CAN SYMBOLIC COMPUTATION AND FORMALIST SYSTEMS ENHANCE MATH EDUCATION WITH ARTIFICIAL INTELLIGENCE?

SEMBOLİK HESAPLAMA VE FORMALİST SİSTEMLER MATEMATİK EĞİTİMİNİ YAPAY ZEKA İLE GELİŞTİREBİLİR Mİ?

Selçuk YAZAR*

Geliş Tarihi: 13.02.2024 *Kabul Tarihi: 07.08.2024*
(Received) *(Accepted)*

ABSTRACT: In recent years, a solution developed using deep learning methods has been used to solve difficult problems in a field. The capability of deep learning models is that they require large and heavily sampled data sets. Computer Algebra Systems developed over time have made significant progress, especially in the field of symbolic mathematics solutions solved by machine learning. It is a persistent problem how appropriate it is to use such formal systems in some aspects of algorithmic decision-making. In this study, a deep learning study in the field of symbolic mathematics and mathematics education is evaluated and the suitability of artificial intelligence applications to formal propositions is discussed. Symbolic computation systems have a strong potential for enhancing math education. Furthermore, within the framework of the Incompleteness Theorem, to show why the construction of a mathematical grammar is not a complete solution for Mathematics education systems.

Key Words: Symbolic Mathematics, Formal Systems, Computer Algebra Systems, Artificial Intelligence, Mathematics Learning.

ÖZ: Son yıllarda, derin öğrenme yöntemleri kullanılarak geliştirilen bir çözüm, bir alandaki zor problemleri çözmek için kullanılmaktadır. Derin öğrenme modellerinin özelliği, büyük ve yoğun örneklenmiş veri setlerine ihtiyaç duymalarıdır. Zaman içerisinde geliştirilen Bilgisayar Cebiri Sistemleri, özellikle makine öğrenmesi ile çözülen sembolik matematik çözümleri alanında önemli ilerlemeler kaydetmiştir. Bu tür biçimsel sistemlerin algoritmik karar vermenin bazı yönlerinde kullanılmasının ne kadar uygun olduğu süregelen bir sorundur. Bu çalışmada özellikle sembolik matematik ve matematik eğitimi alanında yapılan bir derin öğrenme çalışması değerlendirilerek yapay zeka uygulamalarının formal önermelere uygunluğu tartışılmıştır. Sembolik hesaplama sistemleri matematik eğitimini geliştirmek için güçlü bir potansiyele sahiptir. Ayrıca, Eksiklik Teoremi çerçevesinde, matematiksel bir gramer yapısı oluşturmanın Matematik eğitim sistemleri için neden tam bir çözüm olamayacağının gösterilmesi amaçlanmıştır.

Anahtar Kelimeler: Sembolik Matematik, Biçimsel Sistemler, Bilgisayar Cebir Sistemleri, Yapay Zeka, Matematik Öğrenimi.

EXTENDED ABSTRACT

This paper explores the applicability of deep learning methods in formal mathematical systems, specifically focusing on the question of whether deep learning can solve Hilbert's

*Dr. Öğr. Üyesi, Kırklareli Üniversitesi, Mühendislik Fakültesi, selcukyazar@klu.edu.tr, ORCID: 0000-0001-6567-4995.

Entscheidungsproblem. The study investigates the integration of deep learning algorithms and artificial cognitive systems in formal mathematical systems, particularly in the context of symbolic math. It highlights the proposition that "mathematical correctness cannot be expressed in any formalist framework" and discusses the implications of Godel's incompleteness theorem on the applicability of deep learning methods in formal mathematical systems. The paper introduces the concept of explainable AI (XAI) and its relevance in the context of deep learning in formal mathematical systems. The findings suggest that deep learning algorithms, as artificial cognitive systems, can enhance algorithm optimization in formal mathematical systems, including symbolic math.
The study advocates for the integration of artificial cognitive systems and machine learning techniques in mathematics education to improve algorithm optimization. The paper also discusses the development of a grammatical dictionary for mathematical notifications using deep learning methods, as demonstrated in a study by Lample. Overall, the results indicate the potential of deep learning algorithms to address mathematical problems and optimize algorithms in formal mathematical systems, while acknowledging the limitations imposed by Godel's incompleteness theorem.
Considering the mathematical notations used in an approach examined in this study, some defects were observed according to the incompleteness theory. However, the method here also shows us that Gödel's theorem is correct. Because in the data sets here, all possible expressions used in the training of the artificial intelligence model are numbered as sequential proof that can be shown as $\prod_n$. These numbered expressions are sets of t propositions that depend on one or more variables in the model. Within this set (s) there must be a proposition function denoted as the $\boldsymbol{k}^{th}$ which can be expressed as P_k (x). If we set out from the expression number (1), we can express the proposition in the 871^{st} row as the 871^{st} proposition in the form of $P_{871}(x)$, depending on the variable x. So if there will be all the correct axioms in such a set that includes all symbolic expressions, there must also be a negative example of them to be expressed as $\sim P_{871}(x)$. In other words, the statement that there is no evidence for the 871^{st} statement should also be included in this set of propositions. Then, the proposal we expressed with ~P will neither be found in the validation, train, and test sets. If it were found, the deep learning model would learn through errors or contrary mathematical expressions.
Let X_{train} be the training dataset with n samples and m features, and y_{train} be the corresponding labels (0 or 1). Similarly, X_{test} is the test dataset for input. Here, Neural Network architecture is a simple feedforward neural network with a single hidden layer with two parameters. $W^{(1)}$ and $b^{(1)}$ Weights and biases of the hidden layer and $W^{(2)}$ and $b^{(2)}$ Weights and biases of the output layer. Prediction on test set we can apply forward propagation using the updated parameters on X_{test} to obtain predicted probabilities $\hat{y}_{test}$.
For represent the process of training and testing a deep learning algorithm in training process like follows. The training process aims to find the optimal parameters Θ (which includes weights and biases) that minimize the training loss ($\mathcal{L}_{train}$) for the given training data:

$$\Theta^* = argmin_{\Theta}\mathcal{L}_{train}(X_{train}, y_{train}, \Theta) \quad (3)$$

Where $\mathcal{L}_{train}$ quantifies the difference between the model's predictions and the true labels on the training data. In testing process, Let X_{test} be a test sample and D be the same deep learning algorithm. The testing process involves evaluating the model's performance on the test sample. The goal is to predict the label for X_{test} using the learned parameters Θ^* where $\hat{y}_{test} = D(X_{test}, \Theta^*)$. However, if X_{test} is substantially different from the training data distribution or represents an outlier, then the model might not perform well:

$$\hat{y}_{test} \neq y_{true} \tag{4}$$

Where y_{true} is the true label for X_{test}.

The conclusion have reached in this study within the framework of all these views is that the deep learning method has a long way to go in symbolic mathematics and math learning systems. Quantum versions of the artificial neural network studies we have carried out on classical computers will bring us one step closer to the solution of the incompleteness theorem.

1. INTRODUCTION

There are many predictions of artificial intelligence applications. These studies, which are carried out in many different disciplines from autonomous driving, medicine and drug research, natural language processing to artificial vision, allow almost human work to be done by computers and even mobile devices. While these applications of artificial intelligence meet our need for automatas, another expectation continues to develop steadily. Since it was introduced in the 1950s, two different trends have occurred within the concept of artificial intelligence. These are called strong and weak artificial intelligence. Strong artificial intelligence supporters anticipate that artificial learning can one day be developed to a level that reaches human consciousness. The ultimate goal here is to develop a program, algorithm, or mechanical design that endows machines with human-level cognitive capabilities (Searle, 1980).

Deep learning approaches, which are a sub-discipline of artificial intelligence, give very successful results in the research areas that we mentioned above today. Apart from these, another area where the impacts of artificial intelligence are investigated is the research on symbolic mathematics. Computer Algebra Systems (CAS) have been developed to be used in mathematical modeling on computers up to the present day. In these studies, automatic theorem provers and solvers called Boolean Satisfiability Problem (SAT) / Satisfiability Modulo Theories (SMT), which aim to verify the results in logic problems, have been developed. Some of these are currently used for the solutions they need in practice, such as symbolic calculations in graph theory, topology, algebra, or in response to the problems frequently encountered by researchers. Today, systems such as Mathematica (Wolfram) and Magma (Magma), among the most popular of these, perform calculations using algorithms developed for the solution of mathematical problems.

The use of computer algebra systems (CAS) has improved math instruction in a number of ways. CAS give students the opportunity to practice math and programming while fostering the development of computational thinking abilities like abstraction, pattern identification, decomposition, and algorithms. For instanse in paper (Kaneko, Maeda, Hamaguchi, Nozawa & Takato, 2013) The educational impact of resources used in math classes can be improved by using CAS to create accurate 2D visuals that are simple to integrate into superior mathematical writings. The paper proposes a scheme for demonstrating and improving the effect of CAS use in mathematics education by appropriately using the precise 2D-graphics generated by CAS and KETpic. In study (Makishita, 2014) By encouraging students to solve problems rather than just calculate, the use of CAS in the classroom can improve their comprehension of algebra and encourage a more in-depth study of mathematical ideas. The paper discusses the use of Computer Algebra Systems (CAS) in mathematics education and teacher training courses, particularly in the field of geometry. It aims to improve students' ability to use mathematics by incorporating CAS into classroom materials and activities. In addition to making life easier for math teachers, CAS can be utilized as automatic online graders and problem generators for assessments (Heid, Thomas, Zbiek, 2012). Also the paper (Heid, Thomas, Zbiek, 2012) discusses how Computer Algebra Systems (CAS) can improve math education by allowing for new explorations of mathematical concepts, active linking of dynamic representations, engagement with real data, and simulations of real and mathematical relationships. CAS can also enhance students' understanding of algebraic procedures and structures. The paper (Seidametova, 2020) discusses how computer algebra systems, such as Mathematica, can be used in math education to help students practice programming and mathematics skills, and develop computational thinking. It provides an example of using Mathematica for mathematical research on the D(s)-function associated with the Riemann Zeta function. However, it does not specifically mention any improvements related to computer algebra systems (CAS) in mathematics education. These studies focus on exploring the utilization of CAS within the context of mathematics education, particularly within algebra courses. One innovative approach proposed involves employing CAS as "training aids" in algebra, where the system handles arithmetic tasks on behalf of the student, thereby enabling them to concentrate on cultivating their equation-solving skills. The papers also emphasize the implementation of CAS in the introduction of novel concepts, the expansion of procedures, and the exploration of fresh structures in algebra. Another pioneering methodology mentioned is the integration of CAS into the "Application of Mathematics" subject in Japanese high schools, with the objective of enhancing students' mathematical abilities by means of CAS and real-world applications. The papers also address the difficulties encountered by educators in effectively incorporating graphics generated by CAS into printed materials and put forth schemes for demonstrating and enhancing the educational impact of CAS utilization in mathematics education. One paper showcases an instance wherein CAS (Mathematica) is employed for mathematical research pertaining to the D(s)-function associated with the Riemann Zeta function, thereby underscoring the utilization of CAS for problem-solving in mathematics and the provision of research.

The articles propose several potential paths for future investigation and unanswered queries concerning the utilization of computer algebra systems (CAS) in the domain of mathematics education. One suggested direction for further investigation is to delve into the influence of CAS on the instruction and acquisition of algebra within school settings, encompassing its effects on tasks, modes of interaction, and dynamics within the classroom. Moreover, the papers recommend additional research on the implementation of CAS in specific domains, such as geometry, where CAS can prove to be particularly effective for students. There is a call for future research that centers on the utilization of CAS in school algebra, with a particular emphasis on comprehending the theoretical advancement and learning outcomes for students. The papers underscore the necessity for research on how to effectively incorporate CAS graphics into printed educational materials, addressing the challenges faced by teachers in utilizing CAS graphics. One paper suggests that future research should focus on analyzing the causal relationship between the use of CAS graphics and educational outcomes in collegiate mathematics education. Additionally, the documents indicate the necessity for further investigation into the utilization of CAS in diverse mathematical subjects and applications, exploring the potential of CAS in problem-solving, research, and the development of computational thinking abilities. In general, the papers imply that there is still much to learn about the use of CAS in the field of mathematics education, and further research is necessary to fully grasp its impact and potential.

Computer algebra systems can be efficient in understanding mathematics with artificial intelligence. Recent research has shown that machine learning techniques, such as support vector machines, can improve the performance of computer algebra systems by using example problems (England, 2018). In (England, 2018) author discusses the potential for machine learning tools like Support Vector Machines to improve the performance of Computer Algebra Systems, but it does not specifically address the efficiency of computer algebra systems in understanding mathematics with artificial intelligence. Additionally, explainable AI techniques can provide new insights for symbolic computation, inspiring new implementations within computer algebra systems that do not directly rely on AI tools (Pickering, Almajano, England & Cohen, 2024). The use of artificial cognitive systems (ACS) in teaching and learning mathematics has also been proposed, with the understanding that ACS can serve as tools for cognitive reorganization. These ACS tools, such as Matlab and

Mathcad, can aid in the understanding of knowledge produced through their use in mathematics education.

A similar study subject in the field of researching mathematical problems with the help of artificial intelligence is the NP (Nondeterministic Polynomial)-hard and NP-completeness problems. Although the studies in this area are not directly included in mathematics research, the part is related to the question we ask in the title, as they investigate the limit of computability in one aspect. NP time and related problems are a few of the still unsolved problems in computability theory and computational complexity theory. From the symbolic mathematical point of view, this is that the proposition or axioms within the solution of studies and problems require the correct understanding and classification by artificial intelligence. The condition for an artificial intelligence model to be trained according to today's Deep Learning approach have a sufficiently large data set. If a suitable data set can be produced, an answer to the question NP = P can be found with the artificial intelligence approach (Pochart, Jacquot, Mikael, 2022; Ardon,2022). However, the direct relationship between the question we are dealing with in this paper and the NP problems is related to the formation of sets in which these problems will be expressed, as we will mention later.

Previous works in symbolic mathematics were generally research aimed at developing a formal corpus. With the help of these dictionaries, theorem-proving artificial learning algorithms have been developed. In this field, a remarkable study on theorem-proving systems, in which machine learning applications are used in particular, has been published by Müller et al (Müller, Gauthier, Kaliszyk, Kohlhase & Rabe, 2017). In this study, in order to create formal mathematical expressions, they proposed a standard conversion method. The researchers introduced a total of 50,000 alignments with the method they developed. With the approach they applied, they turned the mathematical expressions into URIs compatible with the meta-meta-theory (MMT) format. MMT is a framework for information representation uses formal language such as logic, type theories, ontologies, and set theories, etc.

Translating mathematical expressions into formal and verifiable expressions is one of the first steps in symbolic mathematics studies. Wang et al. converted the mathematical expressions in Latex format into Mizar-Latex expressions using artificial neural networks in their study (Wang, Kaliszyk, Urban, 2018). Mizar is a formal language designed to write advanced mathematical definitions and proofs and is also used as the name of a computer program that can check the proofs in this language. During their studies, the authors managed to convert 1 million aligned Mizar-LATEX pair with an artificial neural network with sequence-to-sequence (seq2seq) architecture with a 65% success rate (Wang, Kaliszyk, Urban, 2018). Thus, they created a symbolic machine-processable corpora that can be used in artificial intelligence applications, and in the studies, we will mention later.

Apart from these approaches, there are also studies on reinforcement learning-based theorem provers. Bansal et al. have developed a reinforcement learning environment called HOList (Bansal, Loos, Rabe, Szegedy & Wilcox, 2019). HOL contains a large solution set of basic mathematical theorems in calculus. In addition, based on this study, they developed a version called DeepHOL that can be adapted to deep learning applications. They used components named goal, tactic, arglist, and neg_arglist in the training sets and in their studies where they obtained a 58% proof success rate.

Another reinforcement learning-based study is the approach applied by Kaliszyk et al. Here, they developed a Monte-Carlo tree search implementation using a prolog-based software library (Kaliszyk, Urban, Michalewski & Olšák, 2018). In their results, they achieved 40% more success than existing mlCoP based systems. Another important paper (Irving, Szegedy, Alemi, Eén, Chollet & Urban, 2016) using deep learning methods and especially based on Kazliyk's work is that of Alemi et al. The researchers used the symbolizations of mathematical expressions using Mizar-Corpus. This is the first study to apply the deep learning approach for theorem proofing systems. They used approximately 52 thousand theorems from Mizar corpus in their data sets. Especially, they developed their models using Convolutional Neural Network layers, which are frequently used today. It is seen that they have achieved a 67% proof success in their results.

Another important study on the mathematical analysis made with convolutional neural networks, including Deep Learning applications, is that of Long et al (Long, Lu, Dong, 2019). In this study, an artificial neural network has been developed to be used in the solution of partial differential equations such as Boltzmann equations and Schrödinger Equation, which are used in solving some physics problems. In the symbolic neural network named SymNet created by the researchers, learning filter functions were used by using a python library that solves more symbolic expressions and is similar to CAS systems. However, as stated, a framework has been developed for solving differential equations using an algorithmic rule and an existing library. In this study, the authors developed an approach that solves these time-dependent differential equations, which are solved based on empirical observations, using sum rules.

The most notable work (Lample & Charton, 2019) on symbolic mathematical calculations was recently published by Lample et al. The different aspect of the mentioned study is that it is a proven-based on a model similar to the Natural Language Processing (NLP) structure, unlike the CAS approaches applied so far. The authors applied the seq2seq method, which is a kind of encoder-decoder approach is frequently used today in language processing in their model. While creating this model, they created a kind of grammatical structure using the mathematical solutions available in the training sets of the model and achieved successful results for symbolic integration and differential equations. Although the approach used in Lample's work which is similar to the automata used by Chomsky to solve grammatical trees (Chomsky & Schützenberger, 1959), their programming and algorithmic approaches are entirely different.

The obtained results prompted us to ask the question in the title of this paper right here. Could be found a grammar consisting of propositions that can be applied to all or some of the mathematical calculations with artificial intelligence?

The aim of this study is to investigate the extent to which machine learning methods employed in symbolic mathematics systems can serve the requirements encountered in mathematics education and mathematical research, within the framework of the incompleteness theorem, taking into account the aforementioned questions and issues.

2. METHODOLGY

Since the first emergence of artificial intelligence applications, studies have been carried out on mathematics/algebra as well as on many subjects of machines. Alan Turing developed and researched the concept of Turing machines in this regard, primarily creating a conceptual system in which natural numbers or other symbolic/formal mathematical expressions can be obtained. However, the subject he investigated is essentially Hilbert's problem known as Entscheidungsproblem (Hilbert & Ackerman, 1928).

The problem known as Entscheidungsproblem can be defined in short as follows; is there a general method that can solve all the problems of mathematics as a principle? In terms of computer science, searching for such a method is equivalent to creating an algorithm. Hilbert's claim was that a formal mathematical system of axioms and methods would contain all provable statements in it. However, in 1931, the German Mathematician Kurt Gödel proved that Hilbert's claim was not valid in the fourth proposition in his doctoral dissertation (Gödel, 1931). It is called the incompleteness theorem. The incompleteness theorem briefly states the following; in a system of propositions that will explain all mathematical propositions (natural numbers, basic arithmetic operations, differential equations, etc.), there must be a statement that is true but cannot be proven. If proof of "x cannot prove" is found in a system of propositions, there will be a contradiction in the system to be created. If there is no contradiction in a system of propositions, the proof of the statement "x cannot prove" cannot be found in this set. Later, Alan Turing proved the correctness of this theorem put forward by Gödel using conceptual Turing machines (Turing, 1936). Generally, this problem is referred to as the "halting problem". The theoretical approach of Turing machines and halting problem requires long-term reading and analysis. But we can summarize it as follows. According to Alan Turing, he thought of a machine that can make calculations as a machine with a tape scanner head and called them a Turing Machine. Mathematical expressions of 0 and 1 values or sequences of algorithms can be found on these bands. These Machines are simply will be able to apply four mathematical operations or rules determined again. The values of 0 and 1 on the band can be changed by moving the band to the right or left. Alan Turing later questioned the verifiability of formal mathematical theories with these machines. With the help of the Universal Turing machine he developed, he discussed whether the steps determined depending on a rule will have an end. He concludes that no program can be written to tell any Turing machine whether it will stop when it runs with the input given, or that a Universal Turing machine cannot be built to indicate it.

3. RESULTS

Gödel's incompleteness theorem has caused us to ask the question in the title on the applicability of deep learning methods in formal mathematical systems. If we take this question further, can an artificial intelligence model be developed that will find solutions to all the problems of mathematics if it is possible to solve symbolic mathematical problems with the methods that will be created based on formal systems?

In order to describe mathematical expressions, an alphabet of symbols is required at first. This alphabet will contain natural numbers, a symbol to separate them, and expressions to be used as variables. Withal, arithmetic operations, parentheses, and logic operators will be required for operation priority. In the study published by Lample, five data sets were used to solve symbolic expressions. As a result, a grammatical dictionary has been developed for mathematical notifications in these data sets.

$$871|sub\ Y'\ mul\ INT-\ 5\sin\ xmul\ INT+\ 5\cos x \qquad (1)$$

For example, the expression (1) used in the validation set shows that an expression has been developed as we mentioned above and that the artificial intelligence model in the study was trained according to this approach. However, the method here also shows us that Gödel's theorem is correct. Because in the data sets here, all possible expressions used in the training of the artificial intelligence model are numbered as sequential proof that can be shown as $\prod_n$. These numbered expressions are sets of *t* propositions that depend on one or more variables in the model. Within this set (s) there must be a proposition function denoted as the ***kth*** which can be expressed as $\mathbf{P}_k$ (x). If we set out from the expression number (1), we can express the proposition in the 871^{st} row as the 871^{st} proposition in the form of $P_{871}(x)$, depending on the variable x. So if there will be all the correct axioms in such a set that includes all symbolic expressions, there must also be a negative example of them to be expressed as $\sim P_{871}(x)$. In other words, the statement that there is no evidence for the 871^{st} statement should also be included in this set of propositions.

Then, the proposal we expressed with ~P will neither be found in the validation, train, and test sets. If it were found, the deep learning model would learn through errors or contrary mathematical expressions.

Let's denote the Entscheidungsproblem as E, and the set of all mathematical statements as S. We can define a hypothetical deep learning algorithm D that attempts to solve E:

$$D : S \rightarrow \{True, False, Undefined\} \quad (2)$$

Where $D(s) = True$ indicates that the deep learning algorithm concludes that statement s is true. $D(s) = False$ indicates that the deep learning algorithm concludes that statement s is false. $D(s) = Undefined$ indicates that the deep learning algorithm cannot make a definitive determination for statement s. However, the key limitation here is that Turing's incompleteness theorem and the undecidability of the Entscheidungsproblem still apply. No matter how advanced the deep learning algorithm is, it cannot escape Gödel's results which show that for any formal system expressive enough to encompass arithmetic, there will always be statements that are undecidable within that system.

This algorithm D is a basic illustration and doesn't reflect the complexities of real-world deep learning architectures. Let's consider a binary classification problem, where we're trying to classify data points into two classes (0 and 1). The algorithm D can be represented as follows:

Lemma 1: Let X_{train} be the training dataset with n samples and m features, and y_{train} be the corresponding labels (0 or 1). Similarly, X_{test} is the test dataset for input. Here, Neural Network architecture is a simple feedforward neural network with a single hidden layer with two parameters. $W^{(1)}$ and $b^{(1)}$ Weights and biases of the hidden layer and $W^{(2)}$ and $b^{(2)}$ Weights and biases of the output layer. Prediction on test set we can apply forward propagation using the updated parameters on X_{test} to obtain predicted probabilities $\hat{y}_{test}$.

Proof 1: For represent the process of training and testing a deep learning algorithm in training process like follows. The training process aims to find the optimal parameters Θ (which includes weights and biases) that minimize the training loss $(\mathcal{L}_{train})$ for the given training data:

$$\Theta^* = argmin_{\Theta}\mathcal{L}_{train}(X_{train}, y_{train}, \Theta) \quad (3)$$

Where $\mathcal{L}_{train}$ quantifies the difference between the model's predictions and the true labels on the training data. In testing process, Let X_{test} be a test sample and D be the same deep learning algorithm. The testing process involves evaluating the model's performance on the test sample. The goal is to predict the label for X_{test} using the learned parameters Θ^* where $\hat{y}_{test} = D(X_{test}, \Theta^*)$. However, if X_{test} is substantially different from the training data distribution or represents an outlier, then the model might not perform well:

$$\hat{y}_{test} \neq y_{true} \tag{4}$$

Where y_{true} is the true label for X_{test}.

It is once again shown that the result we will achieve mechanically (deep learning in this example) confirms the proposition that "mathematical correctness cannot be expressed in any formalist framework" (Rav, 2007).

The fundamental question we address here revolves around whether artificial systems can be developed to solve symbolic mathematical problems, ultimately advancing artificial intelligence applications. If these functions can also be explained with mathematical expressions, will the artificial systems that can do this exactly? J. R. Lucas asked in 1961, is there an algorithm that proves a theorem, similar to the question we asked (Lucas, 1996). According to him if a computer is going to prove a theorem it should use a theorem-proving algorithm. However, there should also be statements that are accepted a priori but cannot be proven to correct in the computer program used. However, the developments in computer technology since the '60s, especially in the field of deep learning, have gone beyond the known algorithmic approaches. The approach of Lamp et al. can also be seen as the research of a theorem-proving model that will prove these theorems. However, according to Lucas, machines cannot perform an uncertain and infinite process. This indicates the necessity of a grammatical dictionary containing symbolic mathematical expressions. The storage and processing powers available today have increased significantly compared to previous years. However, integrating basic expressions into datasets for solving symbolic mathematical problems remains a challenging process, particularly for mathematical theorems that are still undiscovered. The erroneous propositions that we tried to express as ~P above and that we expect to take place in the artificial intelligence model should also be evaluated as this theorem-proving theorem. When we look at it from this point of view, the complexity of the system will increase and, in a sense, a new NP-hard problem will arise.

From another perspective, in the deep learning approach, the learning action ultimately takes place by making use of existing data sets. Since there is a set of activation functions that will give the highest result for each training/test/validate set, how do you find an activation function to prove that the propositions of these sets are false? Then, during the training of the model, it will be necessary to add steps that will show the falsity of the declarations in the list of propositions to be created for all mathematical grammar. Activation functions are linear or non-linear functions used to calculate weighted inputs and bias in artificial neural networks. The type of activation function to be used is decided according to the output of the artificial neural network. An activation function processes the data sent at the input of the network, usually by gradient descent, and then generates an output for the neural network containing the parameters in the input data.

Today, especially in automatic machine learning examples, solutions without human interaction are produced in creating the model and determining the appropriate functions. How will the method of determining the most appropriate parameters for a classification or regression problem, moreover, activation functions, be applied in symbolic mathematical applications under the conditions we question. Today, some solutions of linear and non-linear functions are made with CAS-like algorithms. If the application of determining the most suitable of these functions with artificial intelligence will include unproven theorems, as we mentioned above, the model determination problem will again be linked to the subject of computability. As a result, it is not possible today to obtain a data set that will comply with these principles. Perhaps such proposition sets may be components of an unproven Mandelbrot (Peitgen & Richter, 1986) set.

The studies collectively advocate for the integration of artificial cognitive systems (ACS) and machine learning techniques in mathematics education, providing theoretical justification for their implementation and highlighting their potential to optimize algorithms in symbolic computation. Additionally, they introduce explainable AI (XAI) techniques within this context, offering fresh insights into computer algebra systems. Emphasizing the importance of understanding knowledge generated through these tools, the papers underscore the role of cognitive reorganization and technology in mathematics education. Through case studies and examples, they demonstrate the effectiveness of machine learning and XAI techniques, particularly in variable ordering for cylindrical algebraic decomposition, while also shedding light on their performance compared to existing heuristics and their ability to elucidate decision-making processes in symbolic computation.

Last two years, Large language models, like GPT-3.5 (Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal & Amodei, 2020), are trained using a large amount of text data to learn intricate linguistic patterns. They decrease language modelling loss, which gauges the difference between predicted and real words in phrases, during training by optimizing the parameters. They are exceptional tools for a variety of natural language processing jobs because of this process, which gives them the capacity to produce text that is coherent and contextually relevant. Large language models are anticipated to perform well in testing and generalization on a variety of language tasks outside of their training data. Their performance, like that of deep learning algorithms, may suffer from inputs that dramatically depart from the patterns they have learnt. Outliers, unusual wording, or specialist subjects can make it difficult for them to give precise, appropriate answers. A language model's generation might not agree with an input that is unusual or highly specialized, just as a deep learning model's prediction might not agree with a "false" test sample.

For large language models, the idea of generalization is still essential. Despite their enormous versatility, they are nonetheless susceptible to the drawbacks of generalization. The resemblance of the input to their training data distribution has a significant impact on their ability to generate text. Therefore, when using large language models, making sure the training data is of high quality and taking into account the range of activities they excel at are crucial considerations.

4. CONCLUSION AND DISCUSSION

Symbolic computation systems possess a considerable potential for enhancing the field of mathematics education. These systems, exemplified by MAPLE and muMATH, have the capacity to effectively instruct in various scientific and engineering disciplines through the resolution of algebraic and differential equations, as well as by facilitating precise discussions regarding the outcomes. They afford students the opportunity to acquire knowledge under the guidance of instructors, thereby rendering lectures more captivating and stimulating. The utilization of computer algebra systems on microcomputers and handheld devices is projected to exert a significant influence on the field of high school mathematics education, akin to the impact of electronic pocket calculators. These systems are capable of addressing a vast majority of topics covered in high school and undergraduate mathematics, and future advancements in hardware will expedite their integration into the realm of education. Furthermore, the advantages of employing symbolic computations in the domain of power engineering education are explored, emphasizing the merits of interactive environments for computation, visualization, and modeling.

The Incompleteness theorem, formulated by Godel, posits that there exist veritable mathematical propositions that cannot be demonstrated within a specified axiomatic framework. This signifies that no matter how thorough a formal system is, there will perpetually be veracious propositions that cannot be deduced from the system's axioms. The essence of the Incompleteness theorem coincides with the constraints of constructing a mathematical grammatical structure. The grammatical structure may furnish regulations and principles for mathematical deduction, yet it is incapable of encompassing all conceivable mathematical propositions and their substantiations. Hence, it is unable to ensure a comprehensive comprehension and portrayal of mathematics within an educational framework. The constraints of a mathematical syntax framework become evident when contemplating the extensive range and variety of mathematical assertions and demonstrations. Mathematics, by its inherent essence, is distinguished by its abundance and intricacy, encompassing a plethora of ideas, theories, and occurrences. Endeavoring to encapsulate this abundance within an inflexible syntax framework proves to be a fundamentally challenging endeavor, if not a complete impossibility.

Moreover, the Incompleteness theorem emphasizes the intrinsic incompleteness of any formal system, emphasizing the existence of truths that lie beyond the grasp of formal proof. In the domain of mathematics instruction, it can be inferred that irrespective of the thoroughness with which a curriculum is structured or the accuracy with which concepts are elucidated, there will invariably exist elements of mathematical veracity that elude codification within the confines of a designated framework.

From the perspective afforded by deep learning techniques, the reverberations of Gödel's Incompleteness theorem are palpable and resound within the domain of artificial intelligence and the intricate world of mathematical modeling. Deep learning, which represents a subset of machine learning that takes inspiration from the intricate structure and remarkable functionality of the human brain, places significant reliance on mathematical frameworks and formal systems to effectively handle and navigate through immense quantities of data, thereby facilitating the extraction of meaningful patterns and valuable insights. At its very fundamental core, deep learning functions and operates within the framework of neural networks, which are intricate and complex structures composed of interconnected layers of artificial neurons. These neural networks are trained using vast and extensive datasets, where various mathematical optimization techniques are employed and utilized in order to meticulously adjust and fine-tune the network's parameters, ultimately aiming to minimize and reduce the occurrence of prediction errors. The overwhelming success and triumph of deep learning models is inherently reliant and dependent on their inherent and innate ability to effectively learn and comprehend intricate and convoluted mappings that exist between the inputs and outputs of the system. This unparalleled and exceptional ability enables them to execute a wide range of complex tasks seamlessly and flawlessly, which can range from the highly sophisticated task of image recognition to the intricate and nuanced field of natural language processing, all the while delivering and producing astounding and remarkable levels of accuracy. Much like Gödel effectively and convincingly demonstrated the existence and presence of true mathematical statements that exist beyond the realm and scope of any formal system, deep learning models are inherently and intrinsically bound and constrained by the limitations and confines of the data they are exposed to during the crucial training phase.

Hence, while a mathematical grammatical structure may function as a valuable pedagogical tool, providing a methodical approach to mathematical reasoning and problem-solving, it alone cannot ensure a comprehensive understanding and representation of mathematics within an educational system. Instead, educators must acknowledge the inherent limitations of formal systems and endeavor to cultivate in students a deeper recognition for the innate richness, intricacy, and incompleteness of mathematical truth.

Applying deep learning to symbolic mathematical systems presents numerous obstacles. The aforementioned challenges encompass a range of complexities when it comes to the process of extrapolating findings to cases beyond the originally studied distribution, primarily stemming from the predominant utilization of statistical inference mechanisms as opposed to symbolic reasoning mechanisms. In addition to the points, it is important to note that there exists an insufficiency in the level of transparency and interpretability, both of which pose formidable obstacles to comprehending the intricacies of decision-making processes within intricate mathematical frameworks. Furthermore, effectively integrating domain-specific prior knowledge crucial to these systems proves challenging. Lastly, selecting appropriate model classes is problematic due to the diverse nature of mathematical problems and the limitations of existing deep learning architectures.

Neuro symbolic hybrid systems (Flavio, Alberto, Alessandro, 2023; Zhao, Yang, 2022) have the potential to address the challenges faced in mathematics education when using deep learning methods. These systems integrate concepts from symbolic reasoning, such as computational logic, with deep neural networks. By doing so, these systems enhance the capabilities of deep neural networks by incorporating explainability, the ability to integrate prior knowledge, and modularity. The presented hybrid system, which aims to solve arithmetic problems, showcases the difficulty of applying symbolic reasoning to deep learning models. Deep learning models often struggle to generalize reasoning patterns to cases that are outside their training data. In contrast, the proposed hybrid system utilizes substitution rules in an iterative manner to solve arithmetic expressions and surpasses other models in terms of accuracy.

The conclusion we have reached within the framework of all these views is that the deep learning method has a long way to go in symbolic mathematics. Quantum versions of the artificial neural network studies we have carried out on classical computers will bring us one step closer to the solution of the incompleteness theorem.

Ethical Declaration

In this study, all the rules stated in the "Higher Education Institutions Scientific Research (Türkiye) and Publication Ethics Directive" were followed.

Ethics Committee Approval

The author declare that the research is one of the studies that does not require ethical committee approval.

Conflict of Interest and Funding

No conflict of interest and funding has been declared by the author.

REFERENCES

Ardon, L. (2022). Reinforcement Learning to Solve NP-hard Problems: an Application to the CVRP. arXiv preprint arXiv:2201.05393.

Bansal, K., Loos, S., Rabe, M., Szegedy, C., & Wilcox, S. (2019, May). Holist: An environment for machine learning of higher order logic theorem proving. In International Conference on Machine Learning (pp. 454-463). PMLR.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Chomsky, N., & Schützenberger, M. P. (1959). The algebraic theory of context-free languages. In Studies in Logic and the Foundations of Mathematics (Vol. 26, pp. 118-161). Elsevier.

England, M. (2018). Machine learning for mathematical software. In Mathematical Software–ICMS 2018: 6th International Conference, South Bend, IN, USA, July 24-27, 2018, Proceedings 6 (pp. 165-174). Springer International Publishing. doi: 10.1007/978-3-319-96418-8_20

Flavio, P., Alberto, T., Alessandro, S. (2023). A Hybrid System for Systematic Generalization in Simple Arithmetic Problems. 289-301. doi: 10.48550/arXiv.2306.17249

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für mathematik und physik, 38, 173-198.

Heid, M. K., Thomas, M. O., & Zbiek, R. M. (2012). How might computer algebra systems change the role of algebra in the school curriculum?. In Third international handbook of mathematics education (pp. 597-641). New York, NY: Springer New York. doi: 10.1007/978-1-4614-4684-2_20

Hilbert, D., & Ackerman, W. (1928). Theoretische logik. Julius Springer, Berlin.

Irving, G., Szegedy, C., Alemi, A. A., Eén, N., Chollet, F., & Urban, J. (2016). Deepmath-deep sequence models for premise selection. Advances in neural information processing systems, 29.

Kaliszyk, C., Urban, J., Michalewski, H., & Olšák, M. (2018). Reinforcement learning of theorem proving. Advances in Neural Information Processing Systems, 31.

Kaneko, M., Maeda, Y., Hamaguchi, N., Nozawa, T., & Takato, S. (2013). A Scheme for Demonstrating and Improving the Effect of CAS Use in Mathematics Education. 2013 13th International Conference on Computational Science and Its Applications, 62-71. doi: 10.1109/ICCSA.2013.19

Lample, G., & Charton, F. (2019, September). Deep Learning For Symbolic Mathematics. In International Conference on Learning Representations.

Long, Z., Lu, Y., & Dong, B. (2019). PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. Journal of Computational Physics, 399, 108925. doi: https://doi.org/10.1016/j.jcp.2019.108925.

Lucas, J. (1996). Minds, machines and Gödel: A retrospect. Artificial intelligence: critical concepts, 3, 359-76.

Magma Computational Algebra System. http://magma.maths.usyd.edu.au/magma/

Makishita, H. (2014). Practice with computer algebra systems in mathematics education and teacher training courses. In Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4 (pp. 594-600). Springer Berlin Heidelberg. doi: 10.1007/978-3-662-44199-2_89

Müller, D., Gauthier, T., Kaliszyk, C., Kohlhase, M., & Rabe, F. (2017, June). Classification of alignments between concepts of formal mathematical systems. In International Conference on Intelligent Computer Mathematics (pp. 83-98). Cham: Springer International Publishing.

Peitgen, H. O., & Richter, P. H. (1986). The beauty of fractals: images of complex dynamical systems. Springer Science & Business Media.

Pickering, L., del Río Almajano, T., England, M., & Cohen, K. (2024). Explainable AI Insights for Symbolic Computation: A case study on selecting the variable ordering for cylindrical algebraic decomposition. Journal of Symbolic Computation, 123, 102276.

Pochart, T., Jacquot, P., & Mikael, J. (2022, March). On the challenges of using D-Wave computers to sample Boltzmann Random Variables. In 2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C) (pp. 137-140). IEEE.

Rav, Y. (2007). A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. Philosophia Mathematica, 15(3), 291-320.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424. https://doi.org/10.1017/S0140525X00005756.

Seidametova, Z. (2020). Combining Programming and Mathematics through Computer Simulation Problems. In ICTERI Workshops (Vol. 2732, pp. 869-880).

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. J. of Math, 58(345-363), 5.

Q. Wang, C. Kaliszyk, and J. Urban, "First experiments with neural translation of informal to formal mathematics," 2018: Springer, pp. 255-270.

Wolfram Mathematica. https://www.wolfram.com/mathematica/

Zhao, C., Yang, L. (2022). Symbolic Deep Learning for Structural System Identification. Journal of Structural Engineering-asce, 148(9) doi: 10.1061/(asce)st.1943-541x.0003405