Machine Learning-based Intelligent Formal Reasoning and Proving System

Shengqing Chen*, Xiaojian Huang, Jiaze Fang, Jia Liang

Computer college, Jiaying University, Guangdong, China

*Corresponding author e-mail: hxj_e_mail@163.com

Abstract: The reasoning system can be used in many fields. How to improve reasoning efficiency is the core of the design of system. Through the formal description of formal proof and the regular matching algorithm, after introducing the machine learning algorithm, the system of intelligent formal reasoning and verification has high efficiency. The experimental results show that the system can verify the correctness of propositional logic reasoning and reuse the propositional logical reasoning results, so as to obtain the implicit knowledge in the knowledge base and provide the basic reasoning model for the construction of intelligent system.

1. Introduction

Machine Learning (ML) is a multi-filed inter-discipline, involving many disciplines such as probability theory, statistics, approximation theory, convex analysis and algorithm complexity theory [1] [2]. It focuses on how computers simulate or realize human learning behaviors in order to acquire new knowledge or skills and to reorganize existing knowledge structures to continuously improve their performance. Machine Learning has been applied in various fields, like data mining, computer vision, natural language processing, biometrics, search engines, medical diagnostics, credit card fraud detection, securities market analysis, DNA sequencing, voice and handwriting recognition, etc.

In the past 10 years, Machine Learning has brought a lot of help in such aspects as vehicle driverlessness, practical speech recognition, effective web search, and improvement of human genome awareness. And also, it has been widely used credit decision-making, bioinformatics and power monitoring [3]. It is another important aspect of the AI application to introduce Machine Learning into the research of the logical reasoning system. It plays a very significant role in improving the efficiency of intelligent software.

We have introduced Machine Learning into the process of reasoning, designed intelligent formal reasoning and proving system, realizing its formal reasoning and proving functions. Due to the function of self-learning, the knowledge database will be continuously updated to gradually improve
the efficiency of reasoning and expand the implicit knowledge in the knowledge database, which will provide important help for the construction of various intelligent learning models.

2. Machine Learning in Design

2.1 What is Machine Learning

Machine Learning means a software system improves its performance by gaining experience. The core idea is to let the computer program automatically obtain the accurate judgment and generalization ability with the accumulation of data samples. The essence lies in the data integration, model establishment and algorithm improvement. Throughout the learning process, the most basic conditions are continuous external feedback, external information sources formed in some mode, the obtained external information being processed into "experience" by using algorithm, and the "experience" being stored in an internal database. The database provides the implementation of actions in accordance with the established rules, and the external information obtained in the process of action becomes a new source of feedback, providing new guidance for the next action. The research of Machine Learning should be based on the neural network, statistics like statistical classification and biology, letting a machine simulate the human's learning process. This requires inputting a huge amount of data and learning samples to form "experiences" known to the human, based on which, the relationships between the elements will finally be obtained by repeating division, regression and aggregation, thereby forming the judgment and prediction for similar experiences.

2.2 Some Algorithm

According to the learning way, Machine Learning is mainly divided into three categories: supervised learning, unsupervised learning and semi-supervised learning.

The supervised learning is to generate a model according to the corresponding relationship between the existing input data and output data, map the input into the appropriate output, train the machine to understand the data and the relationship thereof using the tagged data, and output the model of these relationships, so that it can achieve the prediction for the newly-input untagged samples.

The unsupervised learning is to establish a model directly for the input dataset and find out the correlation among data, train the machine using the untagged data to study the structure of the samples and find out the structural relationship among the unknown data, so as to derive a classification model.

The semi-supervised learning is to train the machine using the data with input and output, as well as the data only with input, to learn the structural relationship and output the classification model for prediction. This approach is also the learning algorithm used by the reasoning system in this paper.

2.3 System Structure in Design

A Machine Learning system is mainly composed of three parts: environment, knowledge base and execution module. The environment is the provider of information, which provides necessary information to the learning module of the intelligent system. The learning module uses the obtained information to modify and improve the knowledge base so that the execution module can continuously perform the task more effectively, send the result feedback to the learning module [4] [5]. The following is the relationship between these three components and learning module design.

The learning goal of learning module is to improve the execution module, while another factor affecting the learning module design is the knowledge base. The knowledge base is used to store the
general rules governing the execution module. There are a variety of knowledge expression forms, such as eigenvector, logic statement, semantic network, generative rule and framework. The way to express knowledge is generally considered from the following aspects: ① Whether it can express knowledge clearly; ② Whether it is conducive to knowledge reasoning; ③ Whether it is conducive to modify the knowledge base; and ④ Whether it is conducive to expanding knowledge expression. Without any prior knowledge, the learning system cannot gain knowledge out of thin air, it requires the environment to provide some knowledge as a basis, and then expands and improves such knowledge in order to complete the learning [6].

From the standard of system behavior, "learning is to make the system complete some adaptive changes, so that it can complete the same or similar task more effectively next time"(H.Simon). Within the system, it is reflected as the establishment and improvement of new knowledge structure. Starting from H. Simon’s definition of learning, a simple learning model may be built as shown in Figure 1.

![Figure 1 Learning Model](image)

3. Overview of Reasoning and Proving System

Automated The Proving (ATP) [7] [8], also called Machine Theorem Proving, Mechanization Theorem Proving, etc., is an important branch in the field of artificial intelligence and an inter-discipline of mathematics and computer science [9]. Its research direction is: ① Computer-aided proving, namely machine proving with method adjustment based on the specific. The proving of four-color theorem is one of the examples. It is to give some of the complicated work regarding a particular problem to a computer for resolution by means of computer ② Mechanized proving of system. This proving proposes a general method of dealing with a particular type of problem, which, although cannot solve each type of problem, provides an effective solution to the applicable problems. ③ Logic-based reasoning method, namely the research of proving method. This article is namely to perform the research on the intelligent reasoning system on the basis of the propositional logical reasoning rules.

The most basic and important part of mathematical logic and reasoning is the propositional calculus. The propositional calculus is a method to study how propositions construct more complex propositions through some logical conjunctions and logical reasoning. A propositional formula consists of atomic proposition, conjunction, and truth value (T and F). An atomic proposition is generally expressed with capital letters (such as x1, x2, etc.) and each letter represents a proposition. The atomic propositions compose a compound proposition through the conjunctions which are "∧", "∨", "¬", "→" () and "↔". Refer to the truth value relationships in Figure 2 for the meanings of these conjunctions.
The main reasoning methods for propositional calculus are truth table method, additional premise method and indirect proof method under the natural system. The system mainly uses the additional premise method under the natural system.

For the additional premise method under the natural system, the reasoning structure has the forms of \((A_1 \land A_2 \land \ldots \land A_k) \rightarrow (A \rightarrow B)\). If the conclusion is implicit, the antecedent of the conclusion can be introduced as an antecedent to introduce the consequent. If the reasoning form is implicit, we usually use the additional premise.

Premise: \(A_1, A_2, \ldots, A_k\)

Conclusion: \((A \rightarrow B)\)

Reasoning form: 
\[
(A_1 \land A_2 \land \ldots \land A_k) \rightarrow (A \rightarrow B) \Rightarrow \neg(A_1 \land A_2 \land \ldots \land A_k) \lor (\neg A \rightarrow B) \Rightarrow (\neg(A_1 \land A_2 \land \ldots \land A_k) \lor A) \lor B \Rightarrow (A_1 \land A_2 \land \ldots \land A_k) \rightarrow B.
\]

When we use the additional premise method, a lot of reasoning rules are required. The following are some commonly used formulas:

1. \(A_1 \land A_2 \Rightarrow A_1\)  
   Simplified formula
2. \(A_1 \Rightarrow A_1 \lor A_2\)  
   Additional formula
3. \(\neg A_1 \land (A_1 \lor A_2) \Rightarrow A_2\)  
   Disjunctive syllogism
4. \(A_1 \land (A_1 \rightarrow A_2) \Rightarrow A_2\)  
   Hypothetischer schluss
5. \(\neg A_1 \land (A_2 \rightarrow A_1) \Rightarrow \neg A_2\)  
   Modus tollendo ponens

4. Structure of Reasoning and Proving System

In this paper, the basic learning model in Figure 1 is integrated into the structure of reasoning and proving system. The data is extracted from the knowledge base, the cognition and training of the reasoning process is conducted in the case base where the accurate results have been obtained using the semi-supervised learning algorithm to obtain the truth value which may be used for predicting the same type of reasoning cases. And then, various types of reasoning rule data in the knowledge base is used to find out the correlation among the data in the knowledge base, thus reason the knowledge hidden in the knowledge base through self-learning, and re-store the results into the knowledge base as a new reasoning dataset. This paper is based on propositional logical reasoning system of machine learning, and the system structure is as shown in Figure 3.
The system accepts the inquiry request from the user through the access interface. The user provides the precondition, conclusion and the whole process of formal proving for the deduction. Such process is recorded at the same time, such as the case base and the use of machine learning. And then, the system will call the knowledge case, and the knowledge base can achieve the analysis of the loaded text, as well as the maintenance of the knowledge base itself, including the loading, updating and deletion of knowledge. The reasoning engine realizes the reasoning and proving of the premise, scans each part of the deduction provided by the user one by one and brings it into the knowledge base to parse the correctness of the set so as to judge the correctness of the entire form reasoning. After the reasoning is completed, the reasoning results will be returned to the user by accessing the interface, and reloaded back into the knowledge base for reuse during reasoning later.

Through the semi-supervised learning, the data is continuously read from the case base, the input and output of the reasoning case is regressed and classified through the neural network, thus building the corresponding model, training the machine to recognize the process of reasoning cases so that it can construct the same type of reasoning cases and can predict the result. At the same time, the structure of reasoning rules in semi-supervised learning knowledge base can solve the cluster of rules to find out the correlation, reason the correlation of rules, and record the correlation into the knowledge base, achieving the purpose of self-learning.

5. Implementation and some Results

5.1 Description

Three sets with different properties are created respectively to store the premise collection of deductions, the markMap and the rule collection of tags and conclusions. It describes the reasoning law of each rule through the regular expression of regex. When the initialization is compiled and performed, all the stored knowledge base of rule reasoning laws are stored into the ruleCollection.

In the system access interface, the data about the whole process of deduction provided by the user is stored in the temporary text file judge, and all the proving corollaries are extracted from the judgment text. At each step of reasoning that is formally proved to be the expression x, put the x-expression in the markMap key-value pair with a specific tag and conclusion. Then convert the key to value through the markMap set, for the current step of proving expression x. Form the expression h
that can eventually check the law of matched reasoning. Match the expressions one by one to see if there is a ruleCollection element can be exactly the same. If it exists, mark it as true, otherwise, mark it as false. At the same time, return the false proving directly and stop the operation.

When all matches are over and if all of them are correct, the information that the formal proof of the proposition is correct will be output by printing. At the same time, the premise and conclusion are saved as the reasoning law into the rule knowledge base, so as to provide knowledge information for machine self-learning.

5.2 System Implementation and Test Results

The reasoning system is implemented in Java language, mainly used for the reasoning and testing of the reasoning text of the propositional logic. The test environment is a single CPU host system with Windows operating system and Java virtual machine installed, and the code compilation and running software is Eclipse. A total of 100 reasoning texts of propositional logic were tested, and the reasoning and proving functions were achieved by the system.

The system can correctly reason and prove the reasoning results, and can store the reasoning results into the knowledge base. Through machine learning, the knowledge base is continuously updated and has good execution efficiency.

6. Outlook and Conclusion

The system is a common reasoning platform, and can be expanded to be an intelligent reasoning system with various autonomous learning abilities in the case of an increase of the rules of the field. For example, on the basis of inputting reasoning rule specific to legal logic, the process of legal reasoning of the case is described as an auxiliary handling tool for the judicial personnel. In the aspect of circuit design, it can proof the deduction reasoning of the circuit logic expressions obtained from the truth table and improve the rationality and correctness of the circuit design.

The reasoning system of general platform provides an interface and based on the system, we can refine domain knowledge and express them as supplementary reasoning rules. The system will has strong learning function for specific applications and this prospect is the direction of our further efforts.

Acknowledgements

This article is supported by the following projects
(1) Innovative and Strong School Project of Jiaying College: 315B0128
(2) Teaching Reform Project of Jiaying College: JYJG20170213
(3) Special funds for scientific and technological innovation cultivation project for college students of Jiaying College 2017001

References

[1] Ghahramani, Zoubin. Nature.5/28/2015, Vol.521Issue 7553, p452-459.8p.
[2] YANG BingRu, SONG Wei, XU ZhangYan。Science in China(Series F:Information Sciences)。
[3] Miller-Boling Detlef, Kirchhoff Susanne.Journal of Small Business Management. Apr91, Vol. 29 Issue 2, p83-90. 8p. 2 Diagrams, 1 Chart.
[4] An Lina, Zhang Shijie, Expert system research status and prospects [J], Application Research of Computers, 2007 (12).
[5] Yang Xing, Zhu Dake and Sang Qingbing, Expert system research status and prospects [J], Research on Computer Applications, 2007 (5).

[6] Hou Shifan and Sun Yingjuan. Research on Machine Learning [J] Journal of Changchun Teachers College, 2012 (6).

[7] Guo Siwen, Automatic reasoning program based on user self-added rules [J], Computer Applications and Software, 2007 (9).

[8] Yu Yang, Artificial Intelligence Revolution "Combustion"-Machine Learning, Science & Technology Review, 2016 (7).

[9] Wen kunmei, Lu Zhengding, Wu Jiewen, Li Ruixuan and Sun Xiaolin, Design and Implementation of reasoning system based on description logic-small microcomputer system [J], 2008 (1).