Applications of Computer Algebra – ACA 2024
June 24-28, 2024
Session on *"Machine Intelligence and Symbolic Computation"*

A Framework of Plain Geometric Knowledge Discovery

Dongchen Jiang[1]***, Weifeng Shang***[2]***, Yuting Zhang***[1] [weifengshang@buaa.edu.cn]

[1] School of Information Science and Technology, Beijing Forestry University, Beijing, China
[2] School of Mathematical Sciences, Beihang University, Beijing, China

The transformer-based pre-trained models have played a great role in many fields, in which the large language models, e.g. ChatGPT4.0, have achieved remarkable results. However, it has not achieved satisfactory results in the field of mathematical reasoning and mathematical knowledge generation. In essence, machine learning is the statistical prediction based on the existing empirical data, which is essentially different from automatic reasoning and symbolic computation. Therefore, how to effectively combine the empirical data based prediction with the rule based reasoning or computation has become an issue concerning to us.

By analyzing the process of human reasoning, we find that mathematicians conjecture with the help of experience, and then verifies the relevant properties by using existing properties and theorems. If the properties are successfully verified, they becomes theorems and further become the knowledge of mathematicians. Aiming at the automatic discovery of plane geometry knowledge, we combine the machine learning based methods with the symbolic reasoning based methods to reproduce the process of mathematical knowledge discovery.

Specifically, the knowledge of plane geometry usually includes the following information: the steps to draw a geometric figure, the geometric figure, and the attributes of the figure. Among them, the drawing steps determine the all attributes about the figure, and the specific geometric figure is a concrete representation of the abstract attributes. The attributes of geometric figure include not only the attributes directly contained in the drawing steps, but also the attributes that have not been clearly explained but can be proved by the direct attributes. In this talk, we express these concepts as: drawing steps, geometric figure, direct attributes and indirect attributes.

The purpose of this paper is to explore an automatic method for discovering indirect attributes from given geometric figures. The method includes: 1) using the methods of machine learning to obtain the drawing steps and direct attributes from geometric figures; 2) using the geometric information retrieval algorithm [1] to obtain direct and indirect attributes from geometric figures; 3) using symbolic reasoning tools (or geometric algorithms) to verify the correctness of alternative reasoning attributes; 4) using the GEOTHER [2] module of Epsilon [3] to prove the correctness of the indirect attributes from the direct attributes; 5) combining

all above methods to automatically discover the plane geometry knowledge. More specific work plan of this study is listed as follows:

1. Using a geometric drawing software, such as Geogebra, to draw geometric figures from randomly generated drawing steps; then the dataset of (drawing steps, figure) pairs can be obtained; the set can be used to train a model for generating direct attributes (or drawing steps) from given figures, and this model is called the direct attribute generating model;

2. Using the geometric information retrieval algorithm to obtain direct attributes and the candidates of indirect attributes from the generated geometric figures in 1; the candidate set of indirect attributes is obtained by removing the direct attributes in 1;

3. With the randomly generated direct attributes and corresponding figure in 1, we can try to verify the correctness of the indirect attribute candidates by using the GEOTHER module of Epsilon: if a candidate if verified, it becomes a indirect attribute of the given direct attributes and corresponding figure and will be marked as a knowledge of the plain geometric;

4. By using the dataset of geometric knowledge, i.e. the set of (direct attributes, indirect attributes) pairs, a model for generating indirect attribute candidates can be trained, and we call it as the indirect attribute generating model;

5. For a given geometric figure, the direct attribute generating model is used to generate direct attributes and the indirect attribute generating model is used to generate the indirect attribute candidates, then the geometric module GEOTHER of software Epsilon is used to check the correctness of the candidates. Once a candidate is verified, a new geometric knowledge related with the given figure is obtained automatically.

The above procedure is one implementation about how human beings discover geometric knowledge, i.e. attribute conjecture, attribute verification, and knowledge generation, and it should be total automatic procedure.

Keywords
Geometric information, Image data, Knowledge discovery, Machine learning, Shape recognition, Symbolic computation.

References
[1] D. SONG; D. WANG; X. CHEN, Retrieving geometric information from images: the case of hand-drawn diagrams. *Data Mining and Knowledge Discovery* **31**, 934–971 (2017).
[2] GEOTHER, http://www-polsys.lip6.fr/ wang/GEOTHER/
[3] Epsilon, http://wang.cc4cm.org/epsilon/