Automated Scientific Discovery in Plane Geometry

Tsuyoshi Murata
Department of Computer Science, Institute of Science Tokyo
(formerly: Tokyo Institute of Technology)
W8-59 2-12-1 Ookayama, Meguro, Tokyo 152-8552 Japan
murata@c.titech.ac.jp

The author's attempts in 1990s

Many attempts for automated scientific discovery have been made since 1970s such as MOLGEN (Friedland 1979) and AM (Lenat 1983). Although the computational resources at that time were limited, some researchers pointed out the importance of experimentation in machine discovery (Kulkarni and Simon 1990).

As the attempts for automated scientific discovery in plane geometry, the author proposed DST (Murata et al. 1994) and EXPEDITION (Murata et al. 1996). DST generates figures by changing the angles of a triangle and drawing additional lines, and extracts numerical formulas by observing generated figures. Then the system transforms the formulas by eliminating subproducts that are generated by the addional lines. The transformation is constrained by the number of subproducts in order to avoid the explosive increase of formulas. With little basic knowledge such as the definition of the congruence of triangles and the definition of fundamental trigonometric functions, DST successfully rediscovers many trigonometric formulas and geometrical theorems including the Pythagorean theorem.

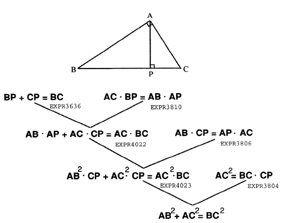

Figure 1: DST

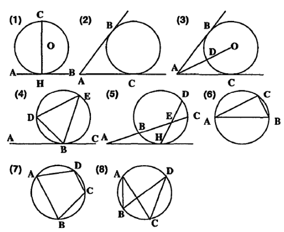

Figure 2: EXPEDITION

EXPEDITION also generates figures by drawing additional lines on a circle. Then the system obtains numerical values of the distances between the pairs of two points and the measures of angles in order to find the candidates of geometrical theorems inductively. In order to make sure that the candidates are not just coincidence, the system generates other similar figures and checks whether the candidates also hold in the new figures. Based on such simple mechanism, EXPEDITION successfully rediscovers geometrical theorems such as Power theorems and Thales' theorem.

Future directions

In order to design effective experiments for obtaining data, counterfactual machine learning (CFML) can be a promising approach. The notion of counterfactual comes from the research community of causal inference. The ability to learn with counterfactuals and generalize to unseen environments is considered as a significant component of general AI. There are some methods for extending CFML to graphs (Guo et al. 2023) (Prado-Romero et al. 2024). Since relations of edges and angles can be represented as graphs, it is expected that such methods can be used for designing experiments for automated scientific discovery.

References

Friedland, P., 1979. Knowledge-based Experiment Design in Molecular Genetics, In Proc. of IJCAI-79, pp.285–287.

Guo, Z., Xiao, T., Aggarwal, C., Liu, H., Wang, S., 2023. Counterfactual Learning on Graphs: A Survey, arXiv:2304.01391, pp.1–50.

Kulkarni, D., and Simon, H. A., 1990. Experimentation in Machine Discovery, In Shrager, J., and Langley, P., eds., Computational Models of Scientific Discovery and Theory Formation, Morgan Kaufmann, pp.255–273.

Lenat, D. B., 1983. The Role of Heuristics in Learning by Discovery: Three Case Studies. In Michalski, R. S.; Carbonell, J. G.; and Mitchell, T. M., eds., Machine Learniing: An Artificial Intelligence Approach, Tioga, pp.243–306.

Murata, T., Mizutani, M., Shimura, M., 1994. A Discovery System for Trigonometric Functions, In Proc. of AAAI-94, pp.645–650.

Murata, T., Shimura, M., 1996. Machine Discovery Based on Numerical Data Generated in Computer Experiments, In Proc. of AAAI-96, pp.737–742.

Prado-Romero, M. A., Prenkaj, B., Stilo, G., Giannotti, F., 2024. A Survey on Graph Counterfactual Explanations: Definitions, Methods, Evaluations, and Research Challenges, ACM Computing Survey, Vol. 56, Issue 7, No. 171, pp.1–31.