Analysis of Scratch Project with Process Data

XiaoLei Wang
Beijing University of Posts and Telecommunications
Beijing 100876, China
Email: 15652964580@163.com

XinMin Huang
Yiwu Research Institute of Education Science
E-mail:2154446@qq.com

Chang Yan
Beijing University of Posts and Telecommunications
Beijing 100876, China
E-mail:937447989@qq.com

Hong Luo
Beijing University of Posts and Telecommunications
Beijing 100876, China
E-mail:luoh@bupt.edu.cn

***Abstract*—Scratch is a visual programming language which has been widely used in the K-12 education. The data generated during the process of student programming can truly reflect behaviors of students during their creation. Analysis of these process data can help us understand the thinking process of students thus does further researches on the students' logical thinking ability. To this end, this paper first designs a novel method of collecting process data using Scratch platform. Then a data analysis method for the complicated process data is proposed. Finally, by comparing with the static data, we verify the correctness of the process data and make analysis from the boys and girls.**

I. Introduction

The process data includes certain actions of students during their creation, such as selection of blocks, cancel of blocks, time of all actions, etc. Through the collection and analysis of process data, we can figure out the students' thinking process, improve their logical thinking ability and formulate personalized training program. In turn, it can be used to cluster and excavate the mass data, obtain the behavior patterns of groups and individuals, and assist teachers to develop individualized training programs. We first analyze the statistical requirements of process data, then find the corresponding code elements in the scratch source code to track, and achieve the automatic collection of process data. According to different situations of students' works, we design the analysis method of process data and verify the correctness of this method by comparing with the static data.

II. PROCESS DATA COLLECTION

When students create projects, they produce a lot of behaviors. According to the principle of computational thinking, meaningless process data such as mouse sliding does not need to be recorded Instead, the type of blocks and the data of connection mode of the blocks need to be recorded. This paper firstly combs the process data record requirements. For example, to analyze the thinking process of students work, we need to record the time of all actions, operations of blocks, etc. Then, we design a process data collection and analysis framework, as shown in Fig. 1. According to the requirements of the process data records, the corresponding Scratch source code is located. When the user submits this work, the process data is uploaded together with the user's work to the backend server. The backend server grades and analyses the submitted work, while extracting static data [1] and process data. The static data includes the number of different blocks and characters. By analyzing process data and static data in a comprehensive way, we can assist teachers to offer each student individual teaching. Based on the programming elements involved in certain work, three types of operations are defined in our work: the operation on the block, the operation on the role and the operation on the backdrop, music and other resource files. To make it convenient for analysis, we design the effective storage structure. For example, the storage structure of block operation is shown in Table I.

TABLE I: Descriptive statistics of collected process data

time	type	op	loc	productions_id	op_type
Dec12 05:41:40	add	op_whenGreenFlag	normal	021a0...c22ad	motion
Dec12 05:41:48	add	op_gotoSpriteOrMouse	normal	021a0...c22ad	motion
Dec12 05:42:02	add	op_yops:	normal	021a0...c22ad	motion

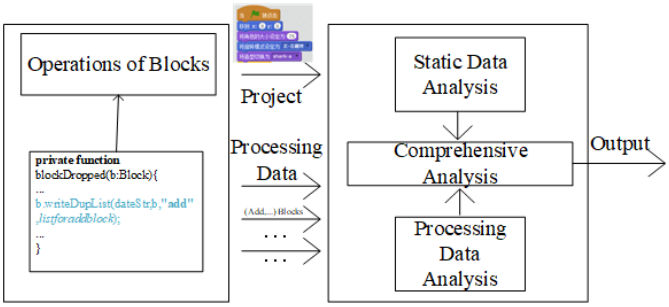

Fig. 1: Process data collection and analysis framework

In Table I, the **time** field indicates the time of operation of the blocks, and the **type** field refers to the type of operation, including addition, deletion, replication, etc., the **op** field represents the name of the blocks, and the **loc** field refers to the connection mode of the blocks, and the **productins_id** refers to the Scratch project number corresponding to the operation, and the **op_type** refers to the category of the instruction module.

III. ANALYSIS METHOD OF PROCESS DATA

To do an effective analysis of the process data, we conduct three coding programs [2] in Yiwu Niansanli Second Primary School. In the use of Scratch platform, there are three creative modes for students as shown in Fig. 2. The first is newly created projects, the second is to improve their own projects, and the third is remix a project. Aiming at the creation of

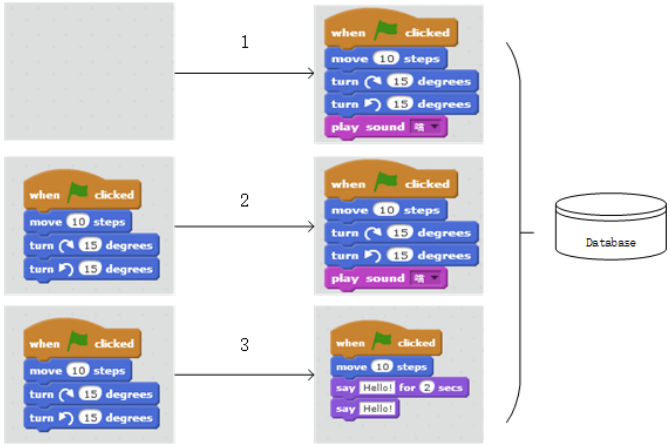

Fig. 2: Three modes of creation

these three types, we store them in different ways. For the first and third, we record all process data directly. For the second, the previous process data is recorded with current process data. The author can also add and cancel block in the process of creating. We define the efficiency in equation 1, while the number of blocks cancelled and the number of new blocks are denoted as del and add, respectively. For a new project, efficiency indicates the proficieny in operation and programming.

$$efficiency = \frac{add - del}{add} \tag{1}$$

IV. FINDINGS

A total of 180 projects are produced in the three coding programs.

A. Correctness Verification

In this paper, static data is compared with the process data in three modes, and the results are shown in Table II:

Since the total number of blocks of most projects is no more than 40, we can conclude that the error is no more than one block, or 2.5 percent is accurate. As shown in Table II, for the new projects, process data is very accurate, and the other two patterns vary slightly. Finally, the overall accuracy of the process data is over 95%. The reason for the deviation is that the work contains unreachable blocks.

B. Analysis of Blocks Usage

In the Scratch, there are 10 types of blocks. The use of different types of blocks varies by the gender of the child, as illustrated in the Fig. 3. From the figure, we can see that the number of boys using all kinds of blocks is higher than that of the girls. For the efficiency of different blocks, boys and girls also show some differences, as shown in Table III: The Table III indicates that male and female students have a huge difference in the use of various blocks. The data block itself is relatively complex, brought about a low use efficiency of

TABLE II: Block operation data storage structure

Error Range	1	2	3	Rate
0-2.5%	130	40	21	96%
2.5-5%	0	4	4	2%
5%-	0	3	1	2%

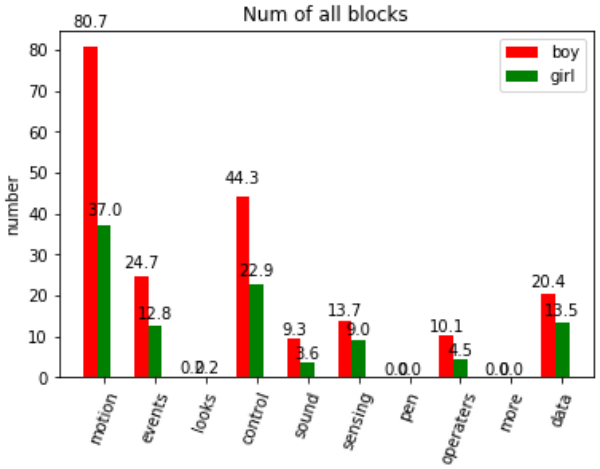

Fig. 3: Three modes of creation

TABLE III: Usage efficiency of male and female students

gender	motion	event	control	sound	sensing	operaters	data
male	0.84	0.81	0.86	0.85	0.87	0.86	0.75
female	0.64	0.66	0.68	0.67	0.71	0.62	0.71

both boys and girls, leading to a small gap. Compared with boys, girls need more training in operator blocks.

V. CONCLUSION

This paper designs a method to collect the process data automatically in Scratch projects creation. We first prove the accuracy of the method by using known static analysis results. After that, we compare the blocks usage of boys and girls, get their efficiency of different blocks. In the future work, we intend to analyze the difference between certain types of projects in process data to analyze programming habits [3].

ACKNOWLEDGMENT

This work is partly supported by the National Natural Science Foundation of China under Grant 61772085, 61672109, 61532012.

REFERENCES

[1] Z. Chang, Y. Sun, T.-Y. Wu, and M. Guizani, "Scratch analysis Tool(SAT): a modern scratch project analysis tool based on ANTLR to assess computational thinking skills," in *IWCMC 2018 Smart Cities and Connected Communities (IWCMC-SmartCity2018)*, Limassol, Cyprus, Jun. 2018.

[2] A. Funke, K. Geldreich, and P. Hubwieser, "Analysis of scratch projects of an introductory programming course for primary school students," in *Global Engineering Education Conference*, 2017.

[3] O. Meerbaum-Salant, M. Armoni, and M. Ben-Ari, "Habits of programming in scratch," in *Sigcse Conference on Innovation and Technology in Computer Science Education, Iticse 2011, Darmstadt, Germany, June*, 2011, pp. 168–172.