ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202009.00125

Small Private Online Judge: A New Tool for Empirical Education Research

Authors: Zhu Yun Chi, Zhao Zuohan, Tong Chengda, Xia Xiaojun, Xia Xiaojun

Date: 2020-09-30T00:00:00+00:00

Abstract

This paper puts forward the concept of Small Private Online Judge (SPOJ). Compared with Massive Open Online Judge (MOOJ), SPOJ has advantages in structured data acquisition of students' virtual behavior for its specific function and tight coupling with the classroom. SPOJ-based empirical education research can be conducted within "Acquisition-Analysis-Application" (3A) Framework. The case study of a SPOJ program clarifies the standard pattern of SPOJ-based 3A research and highlights the emergence of education-intelligence concept. The challenges of SPOJ-based empirical education research and implications of SPOJ are also discussed.

Full Text

Small Private Online Judge: A New Tool for Empirical Education Research

ZHU Yun-chi, School of Biological Sciences & Medical Engineering, Southeast University, Nanjing, China
213170096@seu.edu.cn

TONG Cheng-da, School of Biological Sciences & Medical Engineering, Southeast University, Nanjing, China
213170761@seu.edu.cn

ZHAO Zuo-han, School of Biological Sciences & Medical Engineering, Southeast University, Nanjing, China
213170440@seu.edu.cn

XIA Xiao-jun, School of Biological Sciences & Medical Engineering, Southeast University, Nanjing, China
xxj.rcls@seu.edu.cn

Abstract

This paper introduces the concept of Small Private Online Judge (SPOJ). Compared with Massive Open Online Judge (MOOJ), SPOJ offers superior capabilities for structured data acquisition of students' virtual learning behaviors due to its specialized functions and tight integration with classroom instruction. SPOJ-based empirical education research can be conducted within an "Acquisition-Analysis-Application" (3A) Framework. A case study of an SPOJ implementation clarifies the standard pattern of SPOJ-based 3A research and highlights the emergence of the education-intelligence concept. The challenges of SPOJ-based empirical education research and the broader implications of SPOJ are also discussed.

Keywords: Small Private Online Judge (SPOJ); Empirical Education Research; Virtual Behavior; 3A Framework; Education Intelligence

Background

Online Judge (OJ) is an effective tool for task assignment and learning status assessment widely used in contemporary programming competitions and courses. Through black-box testing in the cloud, it can immediately evaluate code submitted by users and provide corresponding scores and rankings, making it a longstanding component of programming education and algorithm competitions [1].

OJ is not only a high-performance evaluation service platform but also a valuable tool for educational data acquisition [2]. The development of Internet technology has spurred the rise of online learning and associated educational data mining. Compared with traditional classrooms, students' various online learning behaviors are virtualized and structured, making them easier to acquire in a high-throughput manner. Against this backdrop, we propose the role of Small Private Online Judge (SPOJ) in empirical education research related to programming courses.

Basic Concepts of SPOJ

Classification: MOOJ and SPOJ

Similar to online courses, OJ systems can be classified into Massive Open Online Judge (MOOJ) and Small Private Online Judge (SPOJ) according to their different attributes (Table I).

[TABLE I]

Just as SPOC emerged from the booming MOOC movement, SPOJ has developed alongside the evolution and refinement of MOOJ. The launch of early competition-oriented MOOJs such as UVa, POJ, and Codeforces established the foundational development framework and service patterns for OJ systems. Combined with the growing adoption of recent high-quality open-source projects such as HUSTOJ and QDUOJ, current technology enables SPOJs to be rapidly deployed and applied to programming instruction.

Unlike the MOOJs described above, SPOJs are "Small" and "Private" because they are course-oriented OJs built by teachers to provide more efficient programming task assessment for their specific classes. SPOJs are typically deployed on campus networks that are not publicly accessible, thus existing in a numerous-unknown status.

Data Acquisition in SPOJ

The database structure of most OJs includes four main components (Figure 1 [Figure 1: see original paper]):

i) LOG: Log data is the most important data in an OJ' s database. It reflects all user behaviors at specific times on the OJ and is stored in a table named "Solution" or "Status." The basic structure of log data is illustrated in Table II:

[TABLE II]

Compared with other online systems, OJ is more function-specific in that users only submit code and view results there, while actual programming operations are performed locally. Although unable to capture detailed programming behaviors, OJ' s log functions as an efficient programming snapshot system, recording a user' s continuous submissions toward a problem until it is passed or abandoned. Additionally, it provides teachers with access to students' usage habits from the log, such as their usage times and study locations.

ii) PROBLEM: Problem data mainly includes ID, title, description, sample input, sample output, etc. Each problem requires at least one set of test data, which is not usually written to the database. The quantity of problems in SPOJ is much smaller and relatively stable compared with MOOJ, and tends to be adjusted more frequently due to teaching reforms.

iii) TEST: Test refers to the set of time-limited problems released by teachers on SPOJs in the form of exams or assignments for students. The test data includes ID, start and end time, etc. Different from the public problem list of MOOJ, problems on SPOJ are usually posted within tests. To some extent, problems released by teachers are SPOJ' s input signals while logs generated by students' virtual behaviors are its output signals, so teachers tend to utilize tests to keep all signals orderly and bounded.

iv) USER: User data usually includes only the most basic information such as ID and user name when students' detailed information has already been registered on the campus authentication platform.

While sharing similar database structure and scheduling mechanisms with MOOJ, SPOJ offers many advantages in data acquisition. First, SPOJ' s deployer and administrator have a closer relationship (often being the same person), enabling richer types of data to be exported directly from the database instead of requiring web crawlers or APIs. Second, SPOJ is closely integrated with classroom teaching where teachers' assignment release and student task completion are periodic, contributing to more regular data generation. Finally, the scale of SPOJ users is small and stable, and user behavior is much easier to control, which improves SPOJ security and reduces the difficulty of data cleaning.

SPOJ researchers concentrate on log data rather than problem data. For one thing, due to users' elementary programming skills and syllabus regulations limiting the scope of knowledge points, the difficulty and discrimination of course-oriented OJ problems are lower and more uniformly distributed than competition-oriented problems. Therefore, there is less need for SPOJ researchers to focus on problem data to build applications such as learning topic and difficulty level recommendation systems [4]. For another, SPOJ researchers usually have a closer relationship with administrators as well as users, so they may pay more attention to students' virtual behavior patterns that can be mined from log data.

3A Framework Based on SPOJ

SPOJ-based empirical education research fits the "Acquisition-Analysis-Application" (3A) research framework [5] illustrated in Figure 2 [Figure 2: see original paper]. SPOJ serves as the infrastructure supporting these research efforts, where raw data is generated and stored. Researchers can acquire smaller-scale structured data directly from the database and use relatively few computing resources for analysis.

Different from research under the traditional "hypothesis-testing" framework [6], empirical education research under the 3A Framework aims to mine patterns from the unknown to develop applications that solve practical problems. Within the 3A Framework, research can lead to both pedagogical and technological improvements. Pedagogical applications may include student virtual behavior norms, early-warning models fusing OJ data with exam scores, and various education-intelligence reports, while technological applications may involve SPOJ management optimization and classroom-teaching improvement. From this perspective, a complete 3A research project is equivalent to constructing an ETL pipeline connecting SPOJ with applications.

Case Study

Introduction

In 2017, the course leader deployed HUSTOJ on Alibaba Cloud (http://oj.bmeonline.cn) for releasing and assessing programming assignments. Each assignment was released as a test containing 4 to 8 problems with a deadline of 1-2 weeks. User registration was completed by students themselves, and they were permitted to complete tests either off-class or during scheduled on-machine classes. During the three-month period, 76 problems were released on the OJ through 11 tests, generating 10,491 logs from 96 students [7].

In 2018, the teacher again allowed students to register and complete assignments autonomously. The 2017 tests were reused, with only the start dates reset to align with teaching progress; no deadlines were reset. This semester, only 8,493 valid logs were generated on the OJ from 104 enrolled students, representing a year-on-year decrease of 19% [8].

In 2019, Southeast University adopted large-category enrollment and small-class teaching. The original OJ was replaced with QDUOJ (http://www.bmeonline.cn/oj) deployed on the campus LAN, and the user scale was reduced to 26 students from the electronic-information category. OJ registration was changed to unified authentication while supervision of on-machine classes became much stricter. This semester, 12 tests with 68 problems were posted on the OJ, generating 2,609 logs [9].

Acquisition and Analysis

The school's teaching researcher, a data engineer rather than a learning scientist, began exporting and analyzing data from the SPOJ database at the end of 2018.

For dimensionality reduction, most irrelevant variables were removed, including TIME(t) and MEMORY(m), which have long served as criteria for judging algorithm competition problems. These variables either lacked statistical significance (Table III, taking t and m as examples) or had nothing to do with empirical education research.

[TABLE III]

For sample size reduction, when searching for logs unrelated to students in the 2018 data, a large number of subsidiary accounts (alt) were discovered. This pushed the researcher to spend considerable time combining accounts or deleting relevant records directly, causing sample loss beyond prediction. No other abnormal user behavior was found during data cleaning.

The 2017-2018 students' involvement with OJ problems reveals that the number of participants declined exponentially over time (Figure 3 A and B). The gradually increasing difficulty of OJ problems along with classroom-teaching progress may account for intra-test discrimination increase. However, the decline in 2018 appears too large.

The attempt to establish a correlation between students' completion of OJ problems and final course grades did not yield positive results. The researcher defined the AC index according to the ACM judging rule [2] and calculated the Pearson Correlation Coefficient between each student' s AC and their exam scores (MSC&WSC). All results only met the criteria for moderate correlation (Table IV). Such moderate correlation may be caused by the difference between the capability being examined and the capability being trained, as students trained their practical programming skills on the OJ while the final exam focused on theoretical knowledge.

[TABLE IV]

The prominent high AC&MSC correlation in 2017 may be caused by a rare set of students in this school who already had a solid programming foundation before entering university. Their role as teaching assistants during on-machine classes actually multiplied the teaching resources for programming practice in the 2017 course, which consequently increased the MSC and AC correlation to some extent.

Except for 2017, the correlation between AC and MSC is not even as good as WSC, particularly significant in 2018. A survey of the students shows that the majority had received 12 years of exam-oriented education and tended to have formed typical learning habits, leading to the assumption that grasp of theoretical knowledge is more correlated to learning habit than practical skill. Therefore, further observation was made of representation methods for students' learning habits reflected in their virtual behavior on OJ.

Several models were established, among which the Submit Line model (Figure 4, samples from two students in 2017) [2] proved more interpretative. It is determined with linear regression model fitting corrected cumulative submissions against logarithmic date series of a student. This model can reflect both a student' s effort in each test (Kb, slope of Submit Line) and their time arrangement habits for completing them (Stb, R^2 of the fit). The submit-lines in Figure 4 demonstrate that Student 2 is more devoted to OJ problems along with more timely completion than Student 1, which matches their daily submission curves below.

[Figure 4: see original paper]

Application

Findings from the above analysis are applied to two types of practice:

i) Strengthening management of OJ: Contrast between data from the last two years made the course leader realize the necessity for stricter management, leading to the measures mentioned in the introduction (III.A). As a result, the downward trend of the Participation Curve in 2019 (Figure 3 C) becomes much gentler than that of 2018, though not as gentle as that of 2017.

ii) Building education-intelligence (EI) system: Inspired by the business intelligence concept [10], the researcher proposed "education intelligence," which means utilizing advanced data technology to build ETL pipelines within schools to promote the rationality of the 3A-cycle-flow of education data, educational decision-making, and education effectiveness. Guided by this concept, the researcher "employed" one of the three "teaching assistants" (III.B) to transform his analysis methods into software assisting the cleaning, visualization, and analysis of OJ data. The software has been developed since the summer of 2019 and is now open-source under the MIT license [11], representing the completion of the first OJ data ETL pipeline aiding programming teaching.

Discussion

The empirical education research within the 3A Framework reported above has exerted a positive effect on programming teaching practice. It also revealed challenges in SPOJ-based education research:

i) Special behavior influence: SPOJ' s user relationships are usually very close, as users are students learning in the same classroom and even living in the same apartment. Therefore, special behavior of individual users may have a significant impact on the user group, whether positive or negative. While strong measures must be taken to eliminate the effects of negative behavior, positive behavior contributing to a virtuous circle in SPOJ user ecology is more worthy of attention, recording, and research.

ii) Black-box attribute: There is a degree of "deception" in any virtual behavior. A student with a good programming foundation may not take basic problems seriously, while another student who always completes assignments in a timely and efficient manner may actually be a cheater. More data is required to accurately represent such behavior, among which source code data from SPOJ has great potential due to its development systems. Teachers' understanding of student information and behavior-based cluster analysis also help capture such "deceptive" behavior.

iii) Internalization effect: The requirement for strict management of SPOJ makes it inaccessible to the public and restricted in user scale. This guarantees operating efficiency and data quality of SPOJ, but is not conducive to communication between peer teachers or the expansion of education-intelligence. A solution for this dilemma is to establish a public database and open SPOJ data in accordance with a recognized OJ data standard. At present, Xia et al. at SEU has begun to explore this field [12].

iv) Theoretical basis: Despite being application-oriented, SPOJ-based education research still demands theoretical support from learning science, which may explain the more complex cognitive mechanisms and emotional activities behind virtual behaviors on SPOJ, making up for the "dead zone" caused by its black-box attribute from another perspective.

Implications

SPOJ represents the combination of technological innovation and course improvement. It can be defined as "OJ + Classroom," though its implications extend far beyond the classroom. First, education-data-miners on SPOJ may conduct class-useful empirical education research under the guidance of the 3A Framework and take appropriate measures to share their findings and expand education-intelligence. Second, course teachers or leaders can utilize other technologies such as SPOC and "Rain Classroom" to enrich classroom teaching while providing more data for teaching researchers. Last but not least, in today' s information age, disciplines relevant to learning science should attach more importance to student virtual behavior on online platforms such as SPOJ to provide theoretical justification for the analysis and application of education data.

References

[1] Wasik S, Antczak M, Badura J, Laskowski A, Sternal T. A Survey on Online Judge Systems and Their Applications[J]. ACM Computing Surveys. 2018;51(1):31-34. doi:10.1145/3143560.

[2] ZHU Yun-chi, ZHAO Zuo-han, TONG Cheng-da, XIA Xiao-jun. Online Judge Based Programming-practice Data Mining[J]. Journal of Electrical & Electronic Education. 2020,42(1): 94-98+141.

[3] Li Yang. Documents[EB/OL]. https://docs.onlinejudge.me/#/onlinejudge/guide/contest_{rule}_{type}, 2019-09-22.

[4] Wayne Xin Zhao, Wenhui Zhang, Yulan He, Xing Xie, and Ji-Rong Wen. 2017. Automatically Learning Topics and Difficulty Levels of Problems in Online Judge Systems. ACM Trans. Inf. Syst. 36, 3, Article 27 (March 2018), 33 pages. doi:10.1145/3158670

[5] ZHU Yun-chi, ZHAO Zuo-han, TONG Cheng-da, XIA Xiao-jun. A Review of Multimodal Education Evaluation Research under the 3A Framework[J]. Journal of Bio-education: To be published

[6] Dr Robyn Smyth. Introduction to research methods in education -By Keith F Punch[J]. British Journal of Educational Technology,2009,Vol.40(6): 1149-1150

[7] ZHU Yun-chi. Releases NEDB/oj-001[EB/OL]. http://share.bmeonline.cn/nedb/oj-001/releases, 2020-01-30.

[8] ZHU Yun-chi. Releases NEDB/oj-003[EB/OL]. http://share.bmeonline.cn/nedb/oj-003/releases, 2020-01-30.

[9] ZHU Yun-chi. Releases NEDB/oj-005[EB/OL]. http://share.bmeonline.cn/nedb/oj-005/releases, 2020-01-30.

[10] D. J. Power. A Brief History of Decision Support Systems[EB/OL]. http://dssresources.com/history/dsshistory.html, 2007-03-10.

[11] TONG Cheng-da, ZHU Yun-chi. (2020). SPOJ Analyzer (Version 1.0.0.0)[Software]. http://share.bmeonline.cn/admin-and-development/spoj-analyzer/releases

[12] ZHU Yun-chi, XIA Xiao-jun. Release of Virtual Behavior Metadata - Online Judge [EB/OL]. http://www.bmeonline.cn/2020/02/04/%e4%b8%ad%e5%9b%bd%e6%9c%ac%e7%a7%91%e7

ChinaRxiv 中X

1/, 2020-02-04.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.