So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Supporting Group Learning Using Pen Stroke Data Analytics

Daichi YOSHITAKE [a*], Brendan FLANAGAN [b] & Hiroaki OGATA [b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*dai89e@gmail.com

Abstract: Education environments are increasingly becoming digitized, however in some circumstances handwriting input still is an important part of the learning process. In this paper, we propose and implement an educational support system using pen stroke data analysis. Junior high school teachers can perform data analysis themselves with the system. Pen stroke data is collected using a handwritten memo feature in the BookRoll digital learning material reader. Pen stroke data that is collected by the system includes: time, color, x-coordinate and y-coordinate. The system has functions for automatic group formation by using clustering based on the analysis of characteristics of learner pen strokes that are collected from the whole class. It is anticipated that the group formation function can support teachers in quickly creating learning groups in a timely manner based on automated data analysis.

Keywords: learning analytics, pen stroke data, handwriting, educational support, clustering, group learning, group formation

1. Introduction

In recent years, the introduction of information technology into education is progressing. However, there are few cases where learning log data is collected and analyzed to utilize when using digital teaching materials. As a field called learning analytics, research to analyze educational big data is also underway, but discussions have just begun (Hayashi & Hirashima, 2016). Learning analytics is a field that provides educational support by collecting and analyzing educational big data. For example, it is possible to grasp the status of the student's learning efforts from the browsing history of the teaching materials, and recommend appropriate teaching materials.

Meanwhile, group learning is increasingly significant in education since not only cognitive knowledge, but also interpersonal skills such as critical thinking, problem solving and reasoning that count in modern society. Group formation is the most important component since it determines quality of group work (Wessner & Pfister, 2001). Group learning with appropriate groups is found to outperform traditional education (Kyndt et al., 2013). Teachers should create groups based on students' performance and purposes of group learning. On the other hand, teachers usually take a long time on group formation, and it is difficult to create groups in a same class after students solved questions.

In this study, we focus on using pen stroke data obtained when students use a tablet and stylus pen to answer questions with a handwritten memo for group learning. By analyzing the pen stroke data, it is expected that a student's stuck points while answering questions and their understanding level can be estimated. Pen stroke data analytics has already been studied for detecting answer stuck points (Iiyama et al., 2017). However, but group formation using pen stroke data analytics has not been studied. Since features such as stuck points are extracted from the course of the answer by handwriting analysis, it is possible to form a group from a viewpoint that is different from the previous research (Srba, I. & Bielikova, M., 2015). For example, using a heterogeneous group organization it increases the likelihood that students who have different stuck points could teach each other in the same group. In addition, pen stroke pressure is used in some research for educational support (Asai et al., 2012; Takahashi, Imoto, & Yamaguchi, 2015). However, if a system uses pen stroke pressure, students need to use a specified digital pen which supports pen pressure and thus limits the scale to which the analysis can be used. Therefore, in this study, we propose and develop a system that can provide educational support to students by analyzing pen stroke data, and the system is developed as a web application. This allows the system to not be dependent on what devices students use. We provide an interface that allows teachers to analyze pen stroke data themselves. If the data is analyzed by researchers, it cannot be analyzed in real time, and teachers cannot provide immediate support students with results of data analytics. Teachers operate the system themselves to analyze data, so the usefulness of the system in an actual school can be confirmed.

The purposes of this study are to implement a system for group formation using pen stroke data analytics and to confirm the system is ready for being used in actual education sites. In this research, we proposed and implemented a system that allows teachers to analyze pen stroke data by themselves, and conduct a simulation of group formation. We anticipate that the group formation function can support teachers in quickly creating learning groups in a timely manner based on automated data analysis.

2. Digital Learning Material Reading System

2.1 BookRoll

Digitized learning materials are a core part of modern formal education. In addition to serving as a learning material distribution platform, it is also an important source of data for learning analytics into the reading habits of students. The action events of the readers are recorded, such as: turning to the next or previous page, jumping to different pages, memos, comments, bookmarks, and markers indicating parts of the learning materials that are hard to understand or are of importance. The reading behavior of students has previously been used to visualize class preparation and review patterns (Ogata et al., 2017; Yin et al., 2015). The digital textbook system can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

In the present work, the non-proprietary BookRoll digital textbook system was used to serve lecture materials and capture learners reading behavior for analysis (Flanagan & Ogata, 2018). As shown in Figure 1, the user interface supports a variety of functions, such as: navigating to the next or previous page, jumping to any page, marking sections of reading materials in yellow to indicate sections that were not understood, or red for important sections. Memos can also be created at the page level or with a marker to attach it to a specific section of the page. Users can also bookmark pages or use the full text search function to find the information they need to review later. Currently, learning material content can be uploaded to BookRoll in PDF (Portable Document Format) format, and it supports a wide range of devices, including: laptops, tablets, and smartphones, as it can be accessed through a standard web browser.

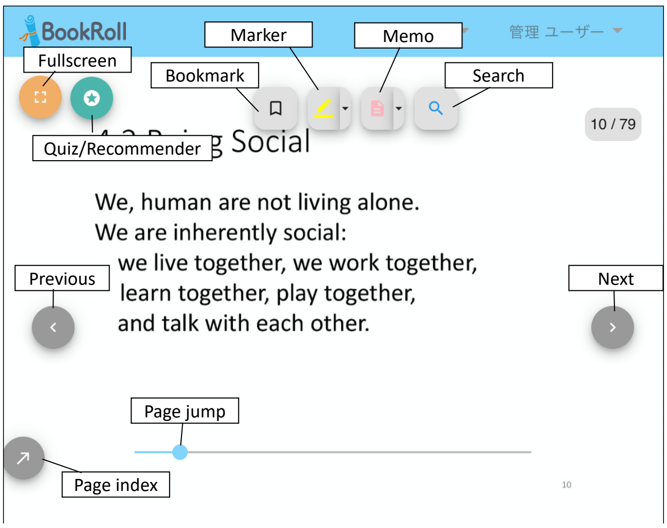

Figure 1. A screenshot of the BookRoll digital learning material.

Reading behavior while using the BookRoll system is send using the xAPI standard for pseudonymized learning event logging and collected in an LRS (Learning Record Store). Table 1 presents a sample of BookRoll's learner behavior logs that have been extracted from an LRS. In the logs there are many types of operations, for example, OPEN means that the student opened the e-book file and NEXT means that he or she clicked the next button to move to the subsequent page.

Table 1. *A sample of events recorded from user interaction with BookRoll.*

Contents id	Memo text	Operation date	Operation name	Page no	User id
EBOOK_341		2018/01/22 18:10	REGIST CONTENTS	0	T1
EBOOK_341		2018/01/23 9:10	OPEN	1	S1
EBOOK_341		2018/01/23 9:20	NEXT	2	S1
EBOOK_341		2018/01/23 9:21	OPEN	1	S2
EBOOK_341	Sample memo	2018/01/23 9:22	ADD MEMO	2	S1
EBOOK_341	rgb(0, 0, 0),1545888614255:56.00522193211488:20.3188491774757,1545888614366:59.45169712793733:20.928109910874298,…	2018/01/23 9:22	ADD HW MEMO	2	S1

2.2 Pen Stroke Input and Data

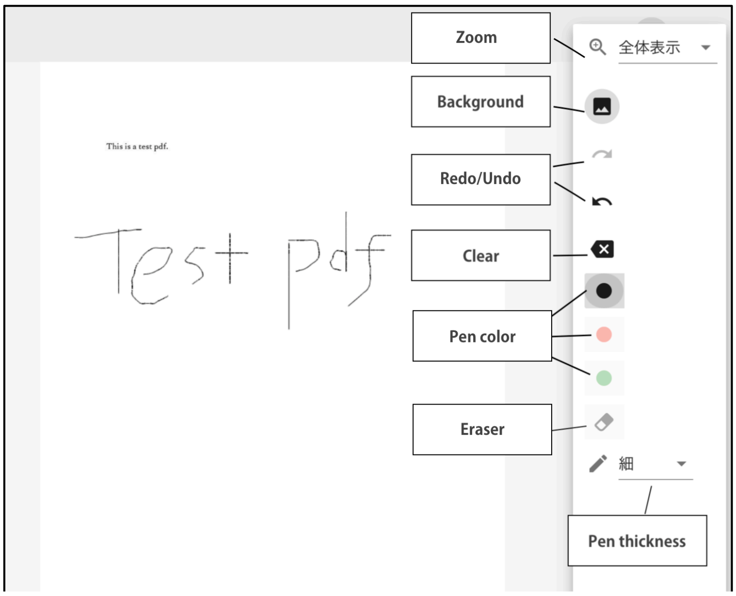

Figure 2. A screenshot of the pen stroke interface in BookRoll.

Figure 2 shows the BookRoll pen stroke input interface. Users can change the pen color and pen thickness, and use the eraser on the right-side menu to remove previously written pen strokes or correct mistakes. Pen stroke data is recorded automatically by BookRoll when a user writes strokes or uses the eraser function. This data is sent to the LRS for collection and analysis using the BookRoll standard xAPI interface as described in the previous section. The pen stroke data format is described below and an example can be found in the last row of Table 2 in the Memo Text column: Pen color, Multiple {UNIX time : x-coordinate : y-coordinate}. The pen color is recorded in rgb format. For example, if the pen color is black, the recorded data will be rgb(0, 0, 0). UNIX time is the time stamp when the data is record, and x-coordinate and y-coordinate express where the pen was at that particular time. UNIX time, and the x-coordinate and y-coordinate are multiple data and additional data is concatenated in a comma separated format, to make up a single pen stroke time series vector.

3. System

We will explain the configuration of the system proposed in this study. This system consists of BookRoll which is a teaching material distribution system, and Analysis Tool and an API for stroke analysis. Students use BookRoll's handwriting memo function to answer questions. The Analysis Tool works with the API (Application Programming Interface) for stroke analysis. The API analyzes the sent stroke data and returns the result. We call this API Stroke API. Stroke API is made with Python and

Flask which is a Python framework. Flask is called a micro framework because it is light, and it works fast. It also takes only a few steps to build an API.

First, students use BookRoll for writing handwritten memos. The pen stroke data from the memo are collected into the LRS automatically. Teachers then select the educational material which they want to analyze on Analysis Tool, and then Analysis Tool extracts the data from the LRS and sends it to the Stroke API. The Stroke API returns the analyzed results, and Analysis Tool then shows the result to the teachers. Teachers can give feedback to students and create learning groups based on the analytics.

Strokes are colored with red when the time interval is more than 2.5 seconds, orange in the case of more than 2.0 seconds, yellow for more than 1.5 seconds, green for more than 1.0 seconds, dark blue when it is more than 0.5 seconds, and black for less than 0.5 seconds. This feature allows teachers to check the student's answer stuck points. Stroke API also supports the eraser function of BookRoll too. The output images from this API also include eraser stroke. In addition, Stroke API extracts the answer time from pen stroke data. Stroke API has a clustering function that clusters handwritten answers based on features using k-means. Given the number of clusters k, k-means classifies the data to be clustered into k clusters and calculates the average of each cluster. The following features are used for clustering: answer time, number of strokes and ratio of six colored strokes and eraser strokes. This system provides a way for teachers to understand student characteristics.

The system has a group formation function based on a result of clustering. The group formation function has 3 types of algorithms: homogeneous, heterogeneous and random algorithms. In the homogeneous algorithm, the system creates groups of students who have same cluster id because it can be said that students who have same cluster id have same features in their answers. In the heterogeneous algorithm, the system creates groups of students who have different cluster id. In the random algorithm, the system just creates groups randomly.

Steps of the homogeneous algorithm is below;

I. Sort students by a number of items in each cluster
II. Add students from the head of students list to a target group until a number of students of a group is satisfied
III. Change the target group to next group
IV. Repeat II and III until finishing formation

Steps of the heterogeneous algorithm is below;

I. Sort students by a number of items in each cluster
II. Add a student from the head of students list to a target group
III. Change the target group to next group
IV. Repeat II and III until finishing formation

Since groups are organized based on learning processes, it is expected to improve the quality of groups, which is important in group learning.

Figure.3 shows the user interface of the group formation function. Teachers set a number of students in 1 group, select a group formation algorithm, then the system creates groups automatically and show a result with student names and their cluster id. Numbers in brackets mean cluster id of those students. Teachers also can download a result as a CSV (Comma Separated Value) file.

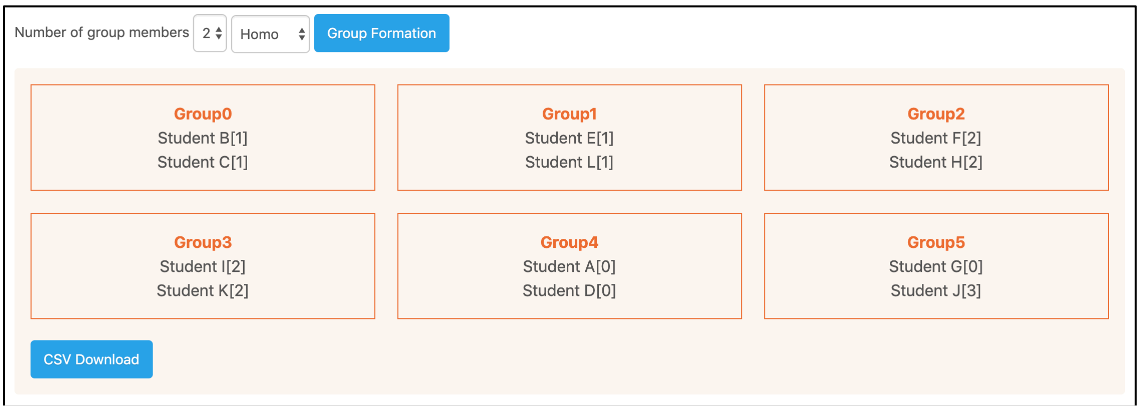

Figure 3. A screenshot of the group formation function.

4. Simulation

We conducted a simulation of group formation using pen stroke data collected from a junior high school mathematics class in Japan. These pen stroke data were collected from 23 students by solving a question about quadratic equation with BookRoll's handwriting memo function. A number of k-means clusters was set to 4. The system created groups based on results of clustering analysis and the number of students belong to a group was set to 4 or 5. We used the homogeneous algorithm and the heterogeneous algorithm to create groups in this simulation.

Table 2 shows the normalized average feature values of each cluster. Features of answer process are confirmed here. For example, students in cluster 2 took eraser much more than students in other clusters. Students in cluster 3 took a long time to write the answer.

Table 2. *Normalized feature values of each clusters.*

Cluster id	Amount of stroke	Eraser	Black stroke	Blue stroke	Green stroke	Yellow stroke	Orange stroke	Red stroke	Answer time
0	0.28	0.12	0.94	0.90	1.00	1.00	0.73	0.13	0.19
1	1.00	1.00	0.41	0.42	0.43	0.26	0.85	0.34	0.57
2	0.40	0.05	0.95	1.00	0.93	0.38	1.00	1.00	0.41
3	0.58	0.05	1.00	0.53	0.76	0.33	0.74	0.70	1.00

5. Results

Table 3. *A result of group formation using the homogeneous algorithm.*

Group 0	Student in cluster[2]	Student in cluster[2]	Student in cluster[2]	Student in cluster[2]	Student in cluster[2]
Group 1	Student in cluster[2]	Student in cluster[2]	Student in cluster[2]	Student in cluster[2]	Student in cluster[1]
Group 2	Student in cluster[1]	Student in cluster[1]	Student in cluster[1]	Student in cluster[1]	Student in cluster[1]
Group 3	Student in cluster[0]	Student in cluster[0]	Student in cluster[0]	Student in cluster[0]	
Group 4	Student in cluster[0]	Student in cluster[3]	Student in cluster[3]	Student in cluster[3]	

Table 4. *A result of group formation using the heterogeneous algorithm.*

Group 0	Student in cluster[0]	Student in cluster[1]	Student in cluster[1]	Student in cluster[2]	Student in cluster[3]
Group 1	Student in cluster[0]	Student in cluster[1]	Student in cluster[2]	Student in cluster[2]	Student in cluster[3]
Group 2	Student in cluster[0]	Student in cluster[1]	Student in cluster[2]	Student in cluster[2]	Student in cluster[3]
Group 3	Student in cluster[0]	Student in cluster[1]	Student in cluster[2]	Student in cluster[2]	
Group 4	Student in cluster[0]	Student in cluster[1]	Student in cluster[2]	Student in cluster[2]	

Table 3 and Table 4 show the results of the group formation simulation. The number in the brackets are the cluster id. We acquired groups of some students who have similar or different types of answers. In Table 3, Groups 0, 2, and 3 were created with only students with the same cluster id. Group 4 has one student who has a different cluster id, but this is inevitable because there is only one student with cluster id 0. There is also only one student with cluster id 1 in group 1. It is possible to place more than one student with the same cluster id if the system exchange students with another group. In this simulation, distances between each cluster were not assumed. If the system considers the distances between clusters, groups would have some students who have different cluster ids, but they still have similar features.

In Table 4, about the heterogenous algorithm, if the number of students in one cluster is large for the number of students with a certain cluster id, a group with all has different ids cannot be created. There is possibility that the heterogeneous algorithm is improved by considering the distance between clusters like the homogeneous algorithm.

6. Conclusions

We proposed and implemented a system for group formation based on a clustering result using pen stroke data. The system creates heterogeneous, homogeneous or random groups. We conducted a simulation of group formation with pen stroke data that was collected from a junior high school mathematics class in Japan. In this paper, we confirmed the system functions as intended by conducting simulation and currently preparations are being made to examine the effectiveness of the system though a formal evaluation experiment in a real classroom.

As future work, we are going to conduct a demonstration experiment to evaluate the system. First students solve mathematics questions using BookRoll's handwriting function. Then, a teacher chooses the heterogeneous or homogeneous algorithm for a purpose and creates groups of students using the system with the chosen algorithm and the random algorithm. We are going to compare performance and feedback of students of groups created by the chosen algorithm and the random algorithm.

Furthermore, considering distances between each cluster for both of the homogeneous algorithm and the heterogeneous algorithm to improve results of group formation including different or same cluster ids.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0 and SPIRITS 2020 of Kyoto University.

References

Asai, H., Nozawa, A., Hamada, S., & Yamana, H. (2012). Detection of learner's trip using online handwritten data. DEIM Forum.

Flanagan, B., Ogata, H. (2018). Learning analytics platform in higher education in japan. *Knowledge Management & E-Learning: An International Journal*, 10(4), 469–484.

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings?. *Educational Research Review*, 10, 133–149, doi: 10.1016/j.edurev.2013.02.002.

Hayashi, Y., & Hirashima, T. (2016). Causal relations from correlations in information-structure oriented approach based educational big data. IPSJ SIG Technical Report (CLE20-1), (pp. 1-6).

Iiyama, M., Nakatsuka, C., Morimura, Y., Hashimoto, A., Murakami, M., & Minoh, M. (2017). Detecting answer stuck point using time intervals of pen strokes. *Transactions of Japanese Society for Information and Systems in Education*, 34(2), 166-171.

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. *In Smart sensors at the IoT frontier*, (pp.327–350). Springer.

Srba, I., Bielikova, M. (2015). Dynamic Group Formation as an Approach to Collaborative Learning Support, in *IEEE Transactions on Learning Technologies*, 8(2), 173-186, doi: 10.1109/TLT.2014.2373374.

Takahashi, S., Imoto, K., Yamaguchi, O. (2015). Estimation of the handwriting behaviors of copying with the online handwriting data. IPSJ SIG Technical Report (CE129-17), (pp. 1-8).

Wessner, M., Pfister, H.-R. (2001). Group formation in computer-supported collaborative learning, in Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work, (pp. 24-31).

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M., Kojima, K. & Ogata, H. (2015). Analyzing the features of learning behaviors of students using e-books. International Conference on Computers in Education (ICCE), (pp. 617–626).