Using Stratified Attribute Tracking (SAT) Diagrams for Learning Analytics

Rwitajit Majumdar

Inter-Disciplinary Program in Educational Technology
Indian Institute of Technology Bombay
rwitajit@iitb.ac.in

Sridhar Iyer

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
sri@iitb.ac.in

Abstract— **We have created a visual representation called Stratified Attribute Tracking (SAT) Diagram to explicate trends that are otherwise implicit in learning analytics data. SAT Diagram is a unified graph that enables tracking individual attribute values in a dataset and stratifying them according to criteria set by the researcher. SAT diagram represents the transition of samples between strata across attributes. In this paper we introduce the SAT diagram and illustrate how to generate, interpret and analyze them. We believe the process of SAT diagram generation would enable exploring deeper research questions on learning data.**

Keywords- Data visualization; Information representations; Learning Analytics; SAT Diagram.

I. INTRODUCTION

With emergence of MOOCs and technology intensive education resources, lot of attention is given to learning analytics (LA) to understand and optimize learning process and its environment [1]. In educational technology (ET) research settings, data is often collected from different sources by applying various educational data mining (EDM) algorithms and typically stored in data tables [2]. Statistical analyses are done on these datasets to provide evidence to support the research objectives [3]. Still it remains cumbersome to draw emerging patterns from the results, that might help in meaningful inference.

Representations and visualizations often come handy for easier interpretation of the organized information. Larkin and Simon's seminal work [4] provides evidence of the importance of external diagrammatic representations to minimize search and computation for implicit elements in the process of problem solving. Card et al. [5] later propose six major ways, which amplify cognition, using visual representations. They increase memory and available resources for processing, reduce the search activities, and by enabling perceptual inference operations help detecting patterns and monitoring. Based on evidence from this distributed cognition literature we developed a diagram to assist visual representation and interpretation of educational datasets.

Our proposed Stratified Attribute Tracking (SAT) diagram, is an unified graph representing the distribution of attribute values of a sample and the transition of that sample from one attribute to another. This representation can be used to explicate patterns from any dataset where information of the cohort as well as the nature of transition of the individuals needs to be visualized.

Typically LA tools still use standard statistical diagrams like scatter plots, line graphs, pie charts and bar charts to report aggregate level and individual sample analysis. We believe the collected data is often rich enough to give further insights to the context that it captured, but it is either lost in the aggregate representation or too cumbersome to interpret by tracking individual data points.

A unified representation of the SAT diagram bridges this gap to assist the LA process. It tracks individual data, then represent it as an aggregate in accordance to criteria set by the researcher. One can then explicate patterns at an intermediate level of zoom of the dataset. The aggregate level overview is also available in the representation.

II. SAT DIAGRAM: STRUCTURE AND INTEPRETATION

Stratified Attribute Tracking Diagram is a unified graph with vertically aligned nodes representing distribution of stratified categories based on attributes of collected data, which are then tracked along the links to the next phase of any activity for a given sample. Between each phase it is a complete bipartite graph. In Figure 1 we give an example of the SAT diagram representation based on the data collected in a Problem Posing (PP) activity [6] and highlight the interpretations that it had.
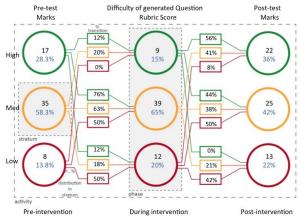

Figure 1. SAT diagram example – The problem posing (PP) activity

SAT diagram can be used to convey the descriptive statistics and also provide assistance for interpretive statistics. A set of vertically aligned circles together represents a *phase*. At each phase a criterion is set to distribute the sample of students in different *strata*. The number of circles in each phase conveys the number of *strata* in that phase. If the strata have an order involved, then the circle representing the higher order stratum is placed on the top. The numerical element in the circle denotes the distribution of the sample in that strata as set by the criteria. The number in the box indicates the

transition ratio, i.e. proportion of the sample in the prior phase that migrates to another stratum in the next phase.

TABLE I.   DESCRIPTORS FOR THE PP PROBLEM SAT DIAGRAM

Phase		Pre-intervention	During intervention	Post intervention
Attribute Noted		Pre-test Marks	Rubric score of difficulty of question	Post-test Marks
Strata & Criteria	High	> 70%	4-5	> 70%
	Medium	> 40% < 70%	2-3	> 40% < 70%
	Low	< 40%	0-1	< 40%

Referring to the PP study, TABLE I. indicates the various attributes that are considered in each phase and the different criteria to set the strata. The resulting distribution of the sample in the three phases is indicated in Figure 1.

The diagram is normally read from left to right, i.e. the distribution of attribute values of pre-phase is on the left and post-phase on the right. Considering transition 1 in the PP study, it is seen that maximum sample population generated questions at a mid level of difficulty. Later in transition 2, 56% of students who generate high difficulty questions also score high in the post-test, while rest 44% become mid-achievers in post-test marks. It is also seen none of them perform low in post-test.

Importantly the whole diagram can be utilized and the particular trend in transition between phases can help in decision making of further action on the data set and look into other attribute recorded for that portion of the sample.

Microsoft Excel® provides the Pivot table function which can assist evaluating the strata distribution and the transition ratio for the collected attribute value data for each phase. The details of the steps are available in the technical report [7].

III.   CASE STUDIES IN EDUCATIONAL RESEARCH

We direct interested readers to refer to 4 published studies that used SAT diagram for exploratory and confirmatory analysis [6]. In the first case study [8] the diagram develops a student behavior model, in the second case [9] it is used to assess the effectiveness of a Large-scale synchronous remote classroom model and in the third case [10] it tracks students' perception data. The PP study [6] is the fourth published result. Currently 5 ET research has used the SAT diagram 8 times to report and interpret their collected data.

IV.   DISCUSSION AND CONCLUSION

SAT diagram facilitates by explicating an external representation of the sample distribution in each phase and assists to track the portion of that sample which migrates to different strata across phases. The flow of the members (mostly student considering ET data set) across the phases can be highlighted from this representation.

Dyckhoff et.al [3] discuss about goals of Learning Analytics. SAT diagram can assist the researcher to achieve many of those goals by providing an external representation. It can track user activities (case study (CS): [8][9]), capture the interaction of students with resources / the interactions among students (CS: [9]), provide an overview (CS: [6]), highlight important aspects of data (CS: [10]), provide different perspectives (CS: [8]), draw the users attention to interesting correlations and provide decision support (CS: [9][10]).

We believe that critical study of the representation techniques in learning analytics might lead to easier and effective analytics. The activity to generate the diagram itself engages the researcher in asking questions on the dataset collected. The cognitive task of setting the criteria of stratification, manipulating the data during pre-processing and the verification of the transition-ratio itself makes the researcher more involved with the data and urges to reflect back on the research questions that it address.   This might help the researcher to improve research skills too.

In this paper we have presented SAT diagram applications to a sub-set of the research question that ET researcher may encounter during applying LA [11]. We propose to examine the set of research questions to which the diagram can be extended.

REFERENCES

[1] L. Ali, M. A. Dragan, G. Gasevi, J. Jovanovi, M. Hatalaa, "Factors influencing beliefs for adoption of a learning analytics tool: An empirical study", *Computers & Education,* vol 62, pp.130–148, 2013.

[2] C. Romero and S. Ventura, "Data mining Educational data mining: A survey from 1995 to 2005", *Expert Systems with Applications,* vol. 33, pp. 135–146, 2007.

[3] A. L. Dyckhoff, V. Lukarov, A. Muslim, M. A. Chatti, and U. Schroeder. 2013. Supporting action research with learning analytics. In *Proc. 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), ACM, New York, NY, USA, pp. 220-229, 2013.

[4] J. H. Larkin, and H. A. Simon, "Why a diagram is (sometimes) worth ten thousand words." *Cognitive science*, vol. 11, no. 1, pp. 65-100, 1987.

[5] K. S. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization, using vision to think.* Morgan Kaufmann, Cal. USA, 1999.

[6] S. Mishra and S. Iyer, "Problem Posing Exercies (PPE): An instructional strategy for learning of complex material in introductory programming courses.", in *Proc IEEE 5th Intnl Conf on Technology for Education (T4E)*, pp.151,158, 18-20 Dec. 2013.

[7] R. Majumdar, "Generating SAT Diagram using Excel's Pivot table.", IIT Bombay, Dept. of Educational Technology, Technical Report (TR-ET-001-2014) Februry 2014. http://www.et.iitb.ac.in/~rwito/overview_excel.pdf

[8] A. Kothiyal, R. Majumdar, S. Murthy and S. Iyer. "Effect of think-pair-share in a large CS1 class: 83% sustained engagement." In *Proc. ACM Intl Computing Education Research Workshop (ICER)*, San Diego, USA, pp. 137-144, August 2013.

[9] J. M. Warriem, S. Murthy and S. Iyer. "Training in-service teachers to do action research in educational technology." in *Proc IEEE 5th Intnl Conf on Technology for Education (T4E)*, pp.192-199, 18-20 Dec. 2013.

[10] R. Majumdar and S. Iyer. "LAMP: A framework for large-scale addressing of muddy points." in *Proc IEEE 5th Intnl Conf on Technology for Education (T4E)*, pp.127-132, 18-20 Dec. 2013.

[11] A. L. Dyckhoff, "Implications for Learning Analytics Tools: A Meta-Analysis of Applied Research Questions", *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 3, pp. 594-601, 2011.