A Machine Learning Approach for Assessment and Prediction of Teamwork Effectiveness in Software Engineering Education

Dragutin Petkovic, Kazunori Okada, Marc Sosnick, Aishwarya Iyer, Shenhaochen Zhu
Department of Computer Science
San Francisco State University
San Francisco, USA
{petkovic, kazokada, msosnick}@sfsu.edu; {aiyer, micz}@mail.sfsu.edu

Rainer Todtenhöfer
Department of Applied Computer Science
Fulda University of Applied Science
Fulda, Germany
rainer.todtenhoefer@informatik.fh-fulda.de

Shihong Huang
Department of Computer Science & Engineering
Florida Atlantic University
Boca Raton, USA
shihong@fau.edu

Abstract - **One of the challenges in effective software engineering (SE) education is the lack of objective assessment methods of how well student teams learn the critically needed teamwork practices, defined as the ability: (i) to learn and effectively apply *SE processes* in a teamwork setting, and (ii) to work as a team to develop satisfactory software (SW) *products*. In addition, there are no effective methods for *predicting* learning effectiveness in order to enable early intervention in the classroom. Most of the current approaches to assess achievement of SE teamwork skills rely solely on qualitative and subjective data taken as surveys at the end of the class and analyzed only with very rudimentary data analysis. In this paper we present a novel approach to address the assessment and prediction of student learning of teamwork effectiveness in software engineering education based on: a) extracting only objective and quantitative student team activity data during their team class project; b) pairing these data with related independent observations and grading of student team effectiveness in SE process and SE product components in order to create "training database"; and c) applying a machine learning (ML) approach, namely random forest classification (RF), to the above training database in order to create ML models, ranked factors and rules that can both explain (e.g. assess) as well as provide prediction of the student teamwork effectiveness. These student team activity data are being collected in joint and already established (since 2006) SE classes at San Francisco State University (SFSU), Florida Atlantic University (FAU) and Fulda University, Germany (Fulda), from approximately 80 students each year, working in about 15 teams, both local and global (with students from multiple schools).**

Keywords-assessment; software engineering teamwork; software engineering education; assessment; machine learning.

I. Introduction and Background

There is now a consensus across industry and academia that to be successful in today's workplace, computer science students must learn practical SE teamwork skills defined as the twofold ability (i) to learn and effectively apply SE processes in teamwork setting and (ii) to work as a team to develop satisfactory software (SW) products which have requested features, and are delivered on budget and on schedule. The need for improved teaching and training in this area is evidenced by statistics on the unacceptably high incidence of failure of industrial SW projects: about ten percent are abandoned, about one third fail, and over half experience cost and schedule overruns [1-5]. The evidence also indicates that these failures stem primarily from failures in communication, organization and teamwork aspects of SE and are not due to the SW technology [1], [4-8].

Most of the current literature on student learning and assessment of SE teamwork skills, while well conceived and developed like CATME [18] and TIDEE [19], or applied on individual instructor basis, relies mostly on qualitative and subjective data from class surveys and instructor observations at the end of the academic term. Due to the subjective and qualitative nature of the collected data, where entries are heavily dependent on the human responder's judgment, these instruments are difficult to use consistently and repetitively. The use of simplistic data analysis methods fail to address complex interactions among team members and the tools they use (e.g. for communication, code management, issue tracking). The absence of objective, quantitative and comprehensive data on student team activities (e.g. team communication dynamics; usage of software development tools) leaves team communication patterns understudied and poorly understood. The fact that the assessments are performed only at the end of the course also precludes early classroom interventions, which are critically important for improving students' learning effectiveness. Sophisticated automated machine learning (ML) techniques that are now regularly applied in bioinformatics, medicine, data mining, marketing, analysis of customer behavior, and even in SE for SW quality assessments (e.g. [9-11]) have not been applied to the acquisition and assessment of SE teamwork skills. The work described in this paper aims to discover new factors that can objectively and quantitatively determine, assess and predict SE student learning teamwork outcomes by applying powerful ML data analysis techniques using only objective and quantitative measures of student team activity.

The authors have been engaged together in joint teaching of SE classes, data collection and some preliminary research since 2006 [12-16]. These SE classes were conducted at SFSU, FAU and Fulda, in a synchronous fashion, using the same team project with the same milestones, with approximately 80 students each year working in about 15 teams. Teams comprised of students only of a particular university (*local* teams) and teams comprised of students from multiple universities (*global* teams).

II. Approach

The approach has several distinct steps (see Figure 1).

A. Step 1: Collection of the data on student team activity

A wide range of data (measures) pertinent to student teamwork activity are collected during the joint SE classes from students while they are actively engaged in intensive team projects. All data are: i) quantitative and objective; ii) related to measurable manifestations of teamwork activity; iii) easy to collect; and iv) amenable to analysis by machine learning methods. All student teams are assigned to develop the same project and fulfill the same five synchronized milestones using the same SE tools (e.g. e-mail server, Bugzilla, SVN) during project development. Teams are formed such that the level of combined expertise and gender mix are approximately equal across teams, in order to factor out the students' skill level from this study. Instructors maintain a log of their regular observations about the teams which are later used for assessment and grading. *Student Activity Measures* (SAM) focus on the activity *of each student* and are obtained by weekly online surveys and analysis of usage of SE tools. These are quantitative measures, such as time used for certain activity, counts of e-mail, incidents, etc., which are either measured by automated tools or easily observed by instructors or students. *Team Activity Measures (TAM)* are computed for each team by combining the *SAM* for the team's members. For example, a SAM datum is the number of commits to the team's source code repository; the corresponding TAM is the average and standard deviation of commits for all the team members. We believe that by focusing only on quantitative variables and combining them at the team level we reduce the influence in reporting error and significantly eliminate subjective bias. To examine different patterns of behavior at different stages of project development, a *time variable* related to each of five project milestones is introduced.

B. Step 2: Creation of ML training database

At the end of the semester, independent evaluators (faculty who do not teach the SE classes) evaluate/grade each student team for achievement of SE teamwork outcomes using: a) the class grading rubrics; b) information from the instructor logs; c) manual evaluation of the developed student SW; and d) final team project demonstration. These grades, one for adherence to the SE *process*) and one for the quality of the team's SE *product* , are categorized as "A - above expectations", "C - at expectations", or "F - below expectations". These grades constitute "decision classes" for the ML algorithm, and are paired with TAM data for each team to construct a *ML training database.*

C. Step3: Applying ML to discover factors that determine and predict student SE teamwork achievement.

The training database will be used as an input to ML training, which will produce a ML classifier that predicts the student team performance based on TAM data, and can assess the effectiveness of TAM measures by evaluating ranked TAM factors. We chose the random forest (RF) [17] ML algorithm for its accuracy, success in many application areas, and its ability to generate simple rules that explain its behavior. We are using open source SW for statistical computing, R [25] for which an easy-to-use RF implementation from [24] is available.

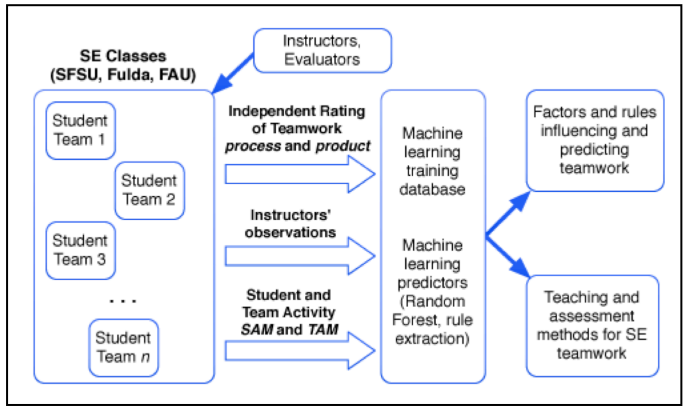

Figure 1. Using ML to assess and predict student teamwork achievement

III. Status and Preliminary Results

We have fully established collaboration, team management and grading methods for the joint SE classes that have been ongoing since 2006 [12-15]. Modified data gathering methods have been in place since Fall 2011 to reflect the new SAM and TAM measurements, when new custom data gathering software started to be used on a new suite of SE tools. These open-source SE tools used by students include: 1) tools for collaboration and communication such as e-mail and wikis; 2) tools for SW development management such as Bugzilla [20] and Subversion **Error! Reference source not found.**; 3) tools for application development such as NetBeans [22]. All team projects are deployed on a server using LAMP (Linux, Apache, MySQL, PHP) stack. Student surveys are administered with LimeSurvey [23]. Tool usage and outcome data is stored in a MySQL database, which is used as training data by the randomForest package [24] for the SW package R [25]. Data analysis has begun, with first results expected by summer 2012. This work is supported in part by NSF TUES grant 1140172.

References

[1] "Standish Group Report: CHAOS Summary 2009." Internet: http://www1.standishgroup.com/newsroom/chaos_2009.php [May 18, 2011].

[2] C. Sauer, C. Cuthbertson, "The State of IT Project Management in the UK 2002-2003." Computer Weekly, 15 April 2003. London.

[3] C. Sauer, A. Gemino, B. Reich,“ The impact of size and volatility on IT project performance,” Communications of the ACM 50(11). November, 2007, pp. 79-84.

[4] C. Jones, Software Engineering best Practices: Lessons from Successful Projects in the Top Companies. McGraw Hill, 2010, ISBN 978-0-07-162161-8.

[5] J. Reel, “Critical success factors in software projects,” IEEE Software 16(3). 1999. pp. 18-23.

[6] R.N. Charette, “Why software fails,” IEEE Spectrum 42(9). September 2005, p. 42.

[7] R. Pressman, Software Engineering: A Practitioner’s Approach, Sixth Edition, McGraw Hill, 2005.

[8] B. Curtis, H. Krasner, N. Iscoe, “A field study of the software design process for large systems.” *Communications of the ACM* 31(11), 1988, pp 1268–1287.

[9] P. Baldi, S. Brunak, Bioinformatics: The Machine Learning Approach, Second Edition (Adaptive Computation and Machine Learning), MIT Press, 2001

[10] P. Perner, ed., “Advances in data mining: applications in medicine, web mining, marketing, image and signal mining,” 6th Industrial Conference on Data Mining, ICDM 2006. Lecture notes in Computer Science, Vol. 4065. Springer. 2006.

[11] E. Zhang, J. Tsai, ed., Machine Learning Applications In Software Engineering. Series on Software Engineering and Knowledge Engineering, Vol. 16. World Scientific, 2005.

[12] E. Petkovic, R. Todtenhöfer, G. Thompson, “Teaching practical software engineering and global software engineering: case study and recommendations,” Proceedings of the 36th ASEE/IEEE Frontiers in Education Conference*,* San Diego, CA, 2006, pp. 19–24.

[13] D. Petkovic, G. Thompson, R. Todtenhöfer, “Teaching practical software engineering and global software engineering: evaluation and comparison.” Proceedings of the Eleventh Annual Conference on Innovation and Technology in Computer Science Education, Bologna, Italy, 2006, pp. 294–298.

[14] D. Petkovic, G. Thompson, R. Todtenhöfer, “Assessment and comparison of local and global SW engineering practices in a classroom setting,” Proceedings of the 13th Annual Conference on Innovation and Technology in Computer Science Education*,* Madrid, Spain, June 2008, pp. 78–82.

[15] G. Thompson, D. Petkovic, S. Huang, R. Todtenhöfer, “Teaching distributed collaborative development techniques in a software engineering class setting.” Integrating FOSS into the Undergraduate Computing Curriculum (FOSS 2009). Chattanooga, 2009.

[16] D. Petkovic, G. Thompson, R. Todtenhöfer, S. Huang, B. Levine, S. Parab, G. Singh, R. Soni, S. Shrestha : “Work in progress: e-TAT : online tool for teamwork and “soft skills” assessment in software engineering education,” Frontiers in Education (FIE) 2010, IEEE. 27-30 Oct. 2010, pp. S1G-1-S1G-3,

[17] L. Breiman, “Random Forests,” Machine Learning 45(1). 2001. pp. 5–32.

[18] “CATME - Comprehensive Assessment for Team-Member Effectiveness.” Intenet: http://www.catme.org [June 6, 2012].

[19] D. Davis, M. Trevisan, P. Daniels, K. Gentili, C. Atman, R. Adams, D. McLean, S. Beyerlein, “ A Model for Transferable Integrated Design Engineering Education”, World Federation of Engineering Organizations , 2003

[20] “Home::Bugzilla::bugzilla.org.” Internet:: http://www.bugzilla.org [June 24, 2012].

[21] “Apache Subverison.” Intenet: http://subversion.apache.org/ [June 24, 2012].

[22] “Welcome to NetBeans.” Internet: http://www.netbeans.org [June 24, 2012]

[23] “LimeSurvey – the free & open source survey tool.” Internet: http://www.limesurvey.org [June 24, 2012].

[24] L. Breiman, A. Cutler , “randomForest: Breiman and Cutler's random forests for classification and regression.” Internet: http://cran.r-project.org/web/packages/randomForest/ [June 24, 2012].

[25] “The R Project for Statisical Computing.” Intenet: http://www.r-project.org/ [June 24, 2012].