Taxonomizing Features and Methods for Identifying At-Risk Students in Computing Courses

Arto Hellas*
University of Helsinki
Helsinki, Finland
arto.hellas@cs.helsinki.fi

Petri Ihantola†
University of Helsinki
Helsinki, Finland
petri.ihantola@helsinki.fi

Andrew Petersen‡
University of Toronto Mississauga
Mississauga, Canada
petersen@cs.toronto.edu

Vangel V. Ajanovski
Saints Cyril and Methodius University
Skopje, Macedonia
vangel.ajanovski@finki.ukim.mk

Mirela Gutica
British
Columbia Institute of Technology
Burnaby, Canada
Mirela_Gutica@bcit.ca

Timo Hynninen
Lappeenranta
University of Technology
Lappeenranta, Finland
Timo.Hynninen@lut.fi

Antti Knutas
Lero,
the Irish Software Research Centre
Limerick, Ireland
antti.knutas@dcu.ie

Juho Leinonen
University of Helsinki
Helsinki, Finland
juho.leinonen@helsinki.fi

Chris Messom
Monash University
Melbourne, Australia
christopher.messom@monash.edu

Soohyun Nam Liao
University of California San Diego
San Diego, USA
snam@eng.ucsd.edu

ABSTRACT

Since computing education began, we have sought to learn why students struggle in computer science and how to identify these at-risk students as early as possible. Due to the increasing availability of instrumented coding tools in introductory CS courses, the amount of direct observational data of student working patterns has increased significantly in the past decade, leading to a flurry of attempts to identify at-risk students using data mining techniques on code artifacts. The goal of this work is to produce a systematic literature review to describe the breadth of work being done on the identification of at-risk students in computing courses. In addition to the review itself, which will summarize key areas of work being completed in the field, we will present a taxonomy (based on data sources, methods, and contexts) to classify work in the area.

CCS CONCEPTS

• **Social and professional topics** → *Computer science education*;

*co-leader
†co-leader
‡co-leader

ITiCSE'18, July 2–4, 2018, Larnaca, Cyprus

ACM ISBN 978-1-4503-5707-4/18/07.
https://doi.org/10.1145/3197091.3205845

KEYWORDS

educational data mining, analytics

ACM Reference Format:
Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Taxonomizing Features and Methods for Identifying At-Risk Students in Computing Courses. In *Proceedings of 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE'18).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3197091.3205845

1 INTRODUCTION

The adoption of instrumented coding tools in introductory CS courses has created an opportunity to directly observe and react to student data. This has lead to increased interest in models that can be used to identify at-risk students in computing courses [3, 4].

This burst of work has lead to a split in the community. Earlier work often relied on the use of student preferences or demographic factors that could be gathered as or before a course began [5]. However, more recent attempts frequently eschew these features, focusing instead on data generated in the course [1, 2]. This working group seeks to connect the various communities – including those outside of computing education – that are supporting the work of identifying at-risk students in computing courses.

The goal of this work is to produce a systematic literature review to describe the breadth of work being done on the identification of at-risk students in computing courses. In addition to the review itself, which will summarize key areas of work being completed in the field, we will present a taxonomy (based on data sources, methods, and contexts) to classify work in the area. We hope the review and accompanying taxonomy will help to connect researchers in this area by identifying clusters of related work being published in different venues and highlighting opportunities for collaboration, integration, and broader dissemination.

REFERENCES

[1] Adam S. Carter, Christopher D. Hundhausen, and Olusola Adesope. 2015. The Normalized Programming State Model: Predicting Student Performance in Computing Courses Based on Programming Behavior. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research (ICER '15)*. ACM, New York, NY, USA, 141–150. https://doi.org/10.1145/2787622.2787710

[2] Matthew C. Jadud. 2006. Methods and Tools for Exploring Novice Compilation Behaviour. In *Proceedings of the Second International Workshop on Computing Education Research (ICER '06)*. ACM, New York, NY, USA, 73–84. https://doi.org/10.1145/1151588.1151600

[3] Alejandro Peña-Ayala. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications* 41, 4 (2014), 1432–1462.

[4] Cristóbal Romero and Sebastián Ventura. 2010. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 6 (2010), 601–618.

[5] Laurie Honour Werth. 1986. Predicting Student Performance in a Beginning Computer Science Class. In *Proceedings of the Seventeenth SIGCSE Technical Symposium on Computer Science Education (SIGCSE '86)*. ACM, New York, NY, USA, 138–143. https://doi.org/10.1145/5600.5701