Re-tain-o-logy: A Comprehensive Model for the Development of Early-Warning and Student Retention Systems

Hossein Hakimzadeh  
Computer Science and Informatics  
Indiana University South Bend  
South Bend, Indiana  
hhakimza@iusb.edu

Cyrus Azarbod  
Information Systems and Technology  
Minnesota State University  
Mankato, Minnesota  
cyrus.azarbod@mnsu.edu

Robert Batzinger  
Computer Science and Informatics  
Indiana University South Bend  
South Bend, Indiana  
rbatzing@iusb.edu

Abstract

In the past several years, student retention has become a major topic of interest in academic circles. The discussion has become even more intricate due to economic and legislative pressures, and administrative directives. As a result, many universities have started to look toward solutions for identifying and reaching out to at-risk students.

In this paper, we will discuss a homegrown early-warning and retention system known as IU-RETAIN which was initiated at Indiana University South Bend in 2007. Contrary to many university IT systems that are developed and operated by the university IT services, the IU-RETAIN system was designed and built by and for faculty and has been in use since 2007. The system has been quite successful in serving approximately 8,700 students per semester. Voluntary faculty participation was at 81% during the spring semester of 2011. More importantly, overall campus retention numbers have significantly improved.

Currently, we are collaborating with faculty at partner institutions to explore and widen the distribution of the methodology as well as the technology employed in IU-RETAIN. In the following sections, we will discuss some of the design principles and technologies employed in building this early-warning and retention system. We will also discuss some of the lessons learned during the process of implementation and operation of the system.

1 INTRODUCTION

In order to help identify at-risk students at Indiana University South Bend, a web-based early warning system has been developed [1]. The system provides a number of portals for faculty, advisors, students and administrators. The data collected and disseminated through these portals are used to quickly identify at-risk students. Once at-risk students are identified, advisors and other staff members focus their resources to help them recover or at least avoid unnecessary failing grades. Although the results are preliminary, early indicators appear to be promising. The system has been quite successful in serving approximately 8,700 students per semester who are taking courses from nearly 1,700 course offerings. The faculty participation started at approximately 30% in the fall of 2007, and has increased nearly every semester to more than 81% during the spring of 2011. During the same period, the Indiana University Institutional Research office reports that the year-to-year retention rate for the campus has improved by approximately 4% [2]. Although numerous efforts contribute to these improvements, one can assert that IU-RETAIN has played an important role, enabling faculty and advisors to identify and reach out to their at-risk students.

2 RETENTION MODEL

IU-RETAIN uses the Re-tain-o-logy model. We define retainology as “...the cross-section of the sciences, methodologies, and technologies used to identify at-risk students, and help them succeed academically.” [3]

The university retention goals are both short-term as well as long-term. The short-term goals are to improve student retention by improving course completion, increase semester-to-semester retention rates, and improve student success by reducing Ds and Fs when possible, and increasing Ws (withdrawals), when appropriate.

These goals are pursued by identifying at-risk students early enough to be able to make a difference, provide support tools to advisors to facilitate effective
intervention, provide reports and meta-analysis to appropriate administrative levels, and to enable timely, accurate and data-driven decision making.

The overall retention model, shown in Figure 1, has matured over the past several years, providing a comprehensive and holistic approach to retention. The model has been implemented in layers, allowing for gradual maturity, and incremental sophistication of the user community. The model was developed first implementing the early-warning and advisor intervention capabilities of the system. The focus was then placed on analysis of data, and assessment of success. More recently, we have turned our attention toward analyzing perceived risk and developing analytical models for predicting risk.

Once at-risk students are identified, efforts to remediate begin immediately. We do so with the help and cooperation of advisors and students themselves.

Risk factors are aggregated and made available to academic advisors, and subsequently, advisors focus their resources on helping students recover or at least avoid unnecessary failing grades. Students are also part of the process. If students are identified by their faculty as having any risk, they are automatically notified and directed to a student portal where they can view the risk status. Once at the student portal, they are also provided with easy access to their faculty, advisor, or other campus resources such as tutoring, counseling, and other services.

Although the results are preliminary, they appear promising. During the past four years, the system has been able to quickly identify students in different risk categories and allow their advisors to intervene in a timely fashion. It has been able to identify courses with the highest number and highest percentage of at-risk students, providing useful information to enable administrators to target resources and take appropriate remedial action.

4 PORTALS

The IU-RETAIN system provides portals to four distinct groups of users (figure 3).

The first group is the faculty, who can use the system, review their class rosters, and provide feedback about their at-risk students, shown in figure 4. The faculty can re-enter the system to update their feedback at any time, and as many times as necessary, for example, after major milestones such as assignments or tests.
The second group of users is the advisors. Academic advisors can start using the system immediately after the initial feedback from faculty has been recorded, typically about two weeks after the start of the semester. Advisors can login as often as they wish and examine the information about their advisees, giving particular attention to those students who are most at-risk, or most likely to be successful via intervention. Once the advisor determines that a student is at-risk, he or she can effortlessly use the built-in communication features to contact their at-risk students, as shown in figure 5.

![My At-Risk Advisees](image)

The third group of users is students themselves. Students receive automatic alerts when they are flagged as having any risk indicators. They are directed to a private and secure student portal where they can observe their overall risk status. Once in the student portal, they are able to conveniently communicate with their professors, advisor, or contact other university services, such as the tutoring center, the counseling center, or the financial aid office.

![Fall 2007 At-Risk Indicators](image)

The fourth group of users is departmental and university administrators. These users are typically interested in reports that show aggregate information about classes, departments, colleges, or the entire university. For example, a department chair may be interested in knowing which of the courses offered by her department have the highest number (or highest percentage) of at-risk students and focus her department’s tutoring, mentoring, and supplemental instructions for those courses. Similarly, the campus and college advising centers may be interested in knowing which freshmen are having difficulty completing their mathematics assignments and target those students individually by calling or sending email encouraging them to come in for tutoring. Finally, university administrators may be interested in tracking student cohorts and comparing them with cohorts from prior academic years.

Although administrative reports can be generated at any time, some administrators may generate these reports on prescribed dates, providing a snapshot of the data that can be compared to previous semesters or years (Figure 6, 7).

![Figure 5](image)

![Figure 6](image)

In addition to the above portals, there is one other portal that is primarily used for system administration and research purposes.

5   THE DATA MODEL

The development of the IU-RETAIN system started with the design of its data model. Given the interest in the topic, many institutions may be interested in developing their own homegrown retention systems. Figure 8 provides a simplified data model which can serve as the starting point for such development. Researchers interested in collaboration in such development are invited to visit the project Web Site [1] or contact the authors directly.

The data model is designed to accommodate a multi-campus university similar to Indiana University or the Minnesota State University system. It allows for role-based access control, campus level customization of interfaces, menus, early-warning questions and other resources.
6 PREDICTION ENGINE

There are many human conditions that are best described in relative terms. Beauty, hunger, and happiness are examples of such conditions. Similarly, being at risk is relative. It is a matter of degree of membership or participation in a “set” of students with risk indicators. Admission specialists can verify that students cannot be classified as at-risk or not at-risk by simply looking at parameters such as prior GPA, SAT score, placement exam score, etc. These parameters simply contribute to our ability to predict the degree of membership in the set.

IU-RETAIN uses a Prediction Engine (P-engine) based on Fuzzy Logic and Set Theory pioneered by Lotfi A. Zadeh in 1965 [5, 6]. Fuzzy sets are sets whose elements have a degree of membership. In classical set theory, elements either belong or don’t belong to the set. In contrast, an element (x) of a fuzzy set (A) has a membership m(x) between 0 and 1. An element (x) is said to be fully included if m(x) = 1. It is said to be not included if m(x) = 0, and it is said to be a fuzzy member if 0 < m(x) < 1.

Prospective students are essentially all fuzzy members of the “set” of at-risk students. The P-engine uses zero to twelve input parameters (based on availability) to calculate the degree to which a student may be at risk prior to the start of the semester.

During the spring semester of 2011, the P-engine was able to successfully predict 63% of the students who eventually received at-risk indicators by their faculty. The system was also able to predict 82% of students who were not likely to be at risk (Figure 9). Collaboration efforts are currently underway to further improve and fine-tune the P-engine [10]. We are also planning to retrospectively apply the P-engine algorithm to data from prior semesters to further validate and refine the algorithm.

7 LESSONS LEARNED

Designing, implementing, and operating a research-based early-warning and retention system has been quite educational. The lessons learned can be categorized into three distinct areas:

- Software engineering
- Usability study and testing
- Operational Challenges

Research projects often have a defined hypothesis and once the central question is answered, the study concludes. We started our research by asking the question:

“Does early-warning feedback by faculty and intervention by advisors improve student retention?”

As indicated in the introduction of this paper, the answer was found to be “yes”. When examining the retention data published by the Indiana University
Institutional Research office, comparing a three year period immediately before the use of IU-RETAIN and three years after the use of the system, one can note approximately 4% improvement in retention rates [2].

Given the above improvement in retention, we are obligated to continue both the operation of the system and our research. The operation must continue because the system is yielding important results that impact student lives. The research must continue in order to refine the retainology model, and answer additional questions such as:

- Can we accurately predict academic risk?
- What remediation measures are most effective in helping at-risk students?
- Can the research results help develop guidelines for admission?
- At what point further efforts become futile, and resources are wasted?
- Should advisor intervention be “more” or “less” intrusive?
- Do students welcome the additional attention and scrutiny resulting from the campus retention efforts? Or, do they view it as an intrusion in their personal or academic life?
- Can the improved retention results be scaled and replicated in other institutions?

8 CONCLUSION

Our overall retention model, shown in Figure 1, has matured over the past several years, providing a comprehensive and holistic approach to retention. Early indications show that the model is able to significantly improve student retention [1]. Longitudinal data provided by Indiana University Institutional Research office, has reported nearly 4% improvement in semester-to-semester and year-to-year retention rates for the campus [2] since the introduction of IU-RETAIN. The re-tain-o-logy model has been implemented in layers, allowing for gradual maturity, and incremental sophistication of the user community. We started by first implementing the early-warning and advisor intervention capabilities of the system. We then focused on analysis of data and assessment of success. More recently, we have turned our attention toward analyzing perceived risk and developing analytical models for predicting risk. Finally, our future goals include the implementation of interface components to course management systems, advising systems, and student self-assessment systems.

9 REFERENCES


[7] MySQL database server is one of the most popular open-source database management systems available at http://www.mysql.com/

[8] PHP is a widely used open-source scripting language for Web development tools available at http://www.php.net/


BIOGRAPHY:

**Dr. Hossein Hakimzadeh** is an associate professor of computer science and director of informatics program at Indiana University South Bend. His research interests include database systems, object-oriented systems, and software engineering.

**Dr. Cyrus Azarbod** is a professor of computer information systems and technology at Minnesota State University at Mankato. His research interests include database security and auditing, active databases, fuzzy active databases, data mining, data warehousing, and software engineering.

**Dr. Robert Batzinger** is a faculty member as well as the Informatics laboratory director at Indiana University South Bend. His research interests include natural language processing: machine-aided document development, automated transcription, word-level back glossing of translations, automatic abstracting and computer-assisted proofreading.