

Detecting At-Risk Students: Empirical Results and Their Theoretical Explanation

Edgar Daniel Rodriguez Velasquez^{1,2},
Olga Kosheleva³, and Vladik Kreinovich⁴

¹Department of Civil Engineering
Universidad de Piura in Peru (UDEP)
Piura, Peru, edgar.rodriguez@udep.pe

²Department of Civil Engineering

³Department of Teacher Education

⁴Department of Computer Science

University of Texas at El Paso

El Paso, Texas 79968, USA

edrodriguezvelasquez@miners.utep.edu,

olgak@utep.edu, vladik@utep.edu

Abstract

In teaching, it is very important to identify, as early as possible, students who may be at risk of failure. Traditionally, two natural criteria are used for this identification: poor grades in previous classes, and poor grades on the first assignments in the current class. Our empirical results show that these criteria do not always work: sometimes a student deemed at-risk by one of these criteria consistently succeeds, and sometimes a student who is not considered at-risk frequently fails. In this paper, we provide a theoretical explanation of our quantitative empirical results, and we use these results to provide recommendations on how to better detect at-risk students.

1 Formulation of the Problem

How do we identify students that need more attention? Instructors and teaching assistants have a limited amount of time to attend to (often a large number of) students. To spend this time wisely, it is important to identify at-risk students, students who need additional attention to succeed. How can we identify these students?

How do we identify at-risk students before class starts: a natural idea. Before the class starts, the only information that we have to identify at-risk students is their past performance. A good indication of the student's past

performance is their average grade in previous classes (in the US, this average grade is called Grade Point Average, GPA for short). If this GPA is low, close to the failure level, it is reasonable to assume that the student is at risk of failure, and additional efforts need to be taken to help these students.

Once the class started, what additional information can we use? Once the course starts, we get the average grade from the first few assignments – and again, these average grades can serve as another indicator of at-risk students:

- if a student has been doing well so far, probably he or she does not need special attention;
- on the other hand, if the student's grades so far on the course's assignments, quizzes, and tests were low, this seems like a good indication that a student needs extra help.

But is this indeed a reasonable strategy? This may sound reasonable, but it is a good idea to check if this strategy indeed works.

What we do in this paper. In our previous research [10], we provided an empirical analysis of the above strategies. In this paper, we recall these results, describe a possible theoretical justification for these empirical results, and provide resulting pedagogical recommendations.

2 Empirical Results: Reminder

What we did. In [10], we studied student performance in several classes, including their starting grade and their final grade in each class.

What we expected. We expected that in most of the cases, both GPA in the previous classes and the average grade for first few assignments would be very good predictors for student's success.

This would mean that all our attention should be concentrated on students with low GPA and/or low performance in the first assignment, and we should not worry that much about other students.

What we found out: somewhat disappointing news. What we found out is that, surprisingly, overall, neither of two usual criteria is a good predictor of the student's success:

- the correlation between the student's success and GPA is low, and
- the correlation between the student's success in a class and the student's average grade for the first few class assignment is also very low.

At first glance, this sounds pessimistic: no way that we can predict the student's success with good accuracy and so, we cannot simply dismiss students who perform well so far as not needing our attention: they may be at risk as well.

What we found out: interesting news. However, a deeper analysis revealed an interesting phenomenon with respect to both correlations.

It turns out that for about a third of all students, their past GPA is a good predictor of the grade in the class. In other words, such students show steady performance, with a low standard deviation of grades from the average.

- Some of these students are straight-A students, maybe with a few B's.
- Some are straight-B students, usually with a few As and Cs.
- Some are straight-C students, with a few Bs and Ds.
- Some are straight-D students, with most grades Ds and fails.

For such students, predicting their success is easy:

- if the student's GPA was passing, we can be reasonably sure that the students will pass this class;
- otherwise, if the student's previous GPA was close to failing, then this student is clearly at-risk, and an additional attention needs to be paid to these students.

Thus is about the third of all students – students with steady performance. For the remaining two thirds of the students, there is no correlation between their GPA and grade in the class.

Similarly, with respect to relation between the preliminary grade in the class and final grade for the same class, there is a similar division:

- For about a third of the students, their initial performance in a class is a good predictor of the student's final grade.
- However, for the remaining two thirds of the students, based on the initial performance, we cannot meaningfully predict their final grade.

Based on two criteria, we have two classes of “predictable” students:

- students whose success in the class can be predicted based on their GPAs, and
- students whose success in the class can be predicted based on their performance in the first weeks of the class.

Each group contains about a third of all the students. These two groups intersect: for some students, their performance in the class can be predicted *both* based on GPA and on their initial performance in the class.

Overall about half of the students are, in this sense, predictable. For the other half, we cannot predict the student's success.

3 Theoretical Explanation of the Empirical Results

Why 1/3: analysis of the problem and the resulting justification. In general, there is a low positive correlation between the final grade for the class and the average grade on the first few assignments. What is a natural way to describe this “low” in precise numerical terms?

Intuitively, the use the word “low” means that we consider two possible values of the positive correlation: low and high. We would like to assign numerical values ℓ and h to these words. The only limitation is that $0 < \ell < h < 1$. Out of many pairs (ℓ, h) that satisfy this inequality, which pair should we choose?

We have no reason to believe that some such pairs are more probable than others. Thus, it is reasonable to assume that all these pairs are equally probable, i.e., that we have a uniform distribution on the set of such pairs. If we want to select a single pair, it is therefore reasonable to select the mean value of (ℓ, h) under this distribution. It is known – see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9] – that this mean corresponds to $\ell = 1/3$ and $h = 2/3$. So, low correlation corresponds to $1/3$.

This explains why in both cases, we get good predictability for $1/3$ of the students.

Why 1/2: analysis of the problem and the resulting justification.

In the previous subsection, we argued that it is reasonable to interpret low correlation between, e.g., the GPA and the grade for a class as correlation of $1/3$. Similarly, we can argue that since the correlation between the GPA and the average of the first grades is also low, we should also describe it by a numerical value of $1/3$.

We now have two events:

- predictability by GPA (we will denote it by A) and
- predictability by the average of the first few grades in a class (we will denote it by B).

We know that the probability $P(A)$ of A is $p = 1/3$, the probability $P(B)$ of B is also $1/3$, and the correlation r between A and B is also equal to $1/3$. What is then the probability $P(A \vee B)$ that for a random selected student, his/her final grade in the class can be predicted based either on the student’s GPA or on the student’s average grade for the few class assignments?

By definition, the correlation r between two random variables x and y is equal to

$$r = \frac{E[x \cdot y] - E[x] \cdot E[y]}{\sigma[x] \cdot \sigma[y]},$$

where $E[\cdot]$ means the mean value and $\sigma[\cdot]$ means standard deviation; see, e.g., [11].

In our case:

- x is a random variable which is equal to 1 if A holds and 0 otherwise, and
- y is a random variable which is equal to 1 if B holds, 0 otherwise.

Since each of the events A and B has probability $p = 1/3$, we can conclude that $E[x] = E[y] = p$. Here, the product of x and y is different from 0 only if both values x and y are different from 0, i.e., when both events A and B occurred, so $E[x \cdot y] = P(A \& B)$.

Here,

$$\sigma^2[x] = E[(x - E[x])^2] = p \cdot (1 - p)^2 + (1 - p) \cdot p^2 = p \cdot (1 - p) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}.$$

Thus, $\sigma[x] = \sigma[y] = \sqrt{2/9}$. Thus, the above formula for r takes the form

$$\frac{1}{3} = \frac{P(A \& B) - 1/9}{\sqrt{2/9} \cdot \sqrt{2/9}} = \frac{P(A \& B) - 1/9}{2/9}.$$

Thus,

$$P(A \& B) = \frac{1}{3} \cdot \frac{2}{9} + \frac{1}{9} = \frac{2}{27} + \frac{3}{27} = \frac{5}{27}.$$

By additivity, we always have

$$P(A \vee B) = P(A) + P(B) - P(A \& B),$$

thus

$$P(A \vee B) = 2 \cdot \frac{1}{3} - \frac{5}{27} = \frac{18}{27} - \frac{5}{27} = \frac{13}{27}.$$

This is very close to the empirical value of $1/2$. Thus, we have indeed provided a theoretical justification for this empirical value.

4 Towards Resulting Recommendations

Before the class starts: analysis of the problem. Before the class starts, we need to check, for each student, whether he/she belongs to the one-third of predictable by GPA. To decide on this, we need to check, e.g., on the standard deviation of all the student's previous grades.

Usually, in the US, we use letter grades. Specifically, the original 0 to 100 numerical grade is transformed into one of the few letter grades: A for excellent, B for good, C for satisfactory, D for probably passing, and F for failing. In computing the GPA, these grades are assigned the following numerical values: A is 4, B is 3, C is 2, D is 1, and F is 0. Of course, if we only use these 0 to 4 grades, we lose a lot of information, so it is better to get and use the original 0-to-100 grades.

For each student, based on his/her previous class grades, we will find the standard deviation.

- If this standard deviation is low, then we can safely predict the student's grade in the course based on his/her GPA.
- Otherwise, we cannot make this prediction.

What threshold should we use? Since about one third is predictable this way, a natural idea is to sort all these standard deviations, and consider those in the lower third predictable-by-GPA.

Thus, we arrive at the following recommendation.

Before the class starts: resulting recommendation. In the beginning, when we do not know anything about the students, any of them can be at-risk.

Before the class starts, to decrease the number of potential at-risk students, we collect, for each student, his/her 0-to-100 grade in all previous classes. Based on these grades, we compute the mean grade and the standard deviation corresponding to this student.

We then sort all the standard deviations, and consider students from the lower third. Those of these lower-third students whose mean grade in previous classes is C or better are clearly not at-risk, so they should be excluded from the list of possible at-risk students.

After the first few grades: analysis of the problem. First, we need to check, for each student from the class, whether this student belongs to the one-third of predictable-by-first-grades. To decide on this, we need to check, e.g., whether there is a correlation in the previous classes between the average grade on the first several assignments in each class and the final grade for the same class.

For this purpose, we need to know not only the student's 0-to-100 final grades in all previous classes, we also need to know, for each of these previous classes, the student's average of the first 0-to-100 grades in this class.

Based on this information, we can compute, for each student, the correlation between the average of the first few grades and the final grade for the corresponding class.

- If this standard correlation is high, then we can safely predict the student's final grade in the course based on his/her average grade from the first few assignments.
- Otherwise, we cannot make this prediction.

What threshold should we use? Since about one third is predictable this way, a natural idea is to sort all these correlations, and consider those in the upper third predictable-by-first-grades.

Thus, we arrive at the following recommendation.

After the first few grades: resulting recommendation. Before the class starts, for each student and for each class that this student attended, we collect, in addition to the student's 0-to-100 final grade for this class, an average of the 0-to-100 grades for the first several assignments. Based on this information,

we compute, for each student, the correlation between the final grade and the average-of-first grades.

We then sort all the resulting correlations, and consider students from the upper third. Those of these upper-third students whose average on the first few grades in this class is C or better are clearly not at-risk, so they should be excluded from the list of possible at-risk students.

The resulting list – consisting of slightly more than half students – contains everyone who can potentially be at risk.

Acknowledgments

This work was partially supported by the US National Science Foundation via grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

- [1] M. Ahsanullah, V. B. Nevzorov, and M. Shakil, *An Introduction to Order Statistics*, Atlantis Press, Paris, 2013.
- [2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, Society of Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, 2008.
- [3] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley, New York, 2003.
- [4] O. Kosheleva, V. Kreinovich, J. Lorkowski, and M. Osegueda, “How to transform partial order between degrees into numerical values”, *Proceedings of the 2016 IEEE International Conferences on Systems, Man, and Cybernetics SMC’2016*, Budapest, Hungary, October 9–12, 2016.
- [5] O. Kosheleva, V. Kreinovich, M. Osegueda Escobar, and K. Kato, “Towards the most robust way of assigning numerical degrees to ordered labels, with possible applications to dark matter and dark energy”, *Proceedings of the 2016 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS’2016*, El Paso, Texas, October 31 – November 4, 2016.
- [6] J. Lorkowski and V. Kreinovich, “Interval and symmetry approaches to uncertainty – pioneered by Wiener – help explain seemingly irrational human behavior: a case study”, *Proceedings of the 2014 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS’2014*, Boston, Massachusetts, June 24–26, 2014.
- [7] J. Lorkowski and V. Kreinovich, “Likert-type fuzzy uncertainty from a traditional decision making viewpoint: how symmetry helps explain human decision making (including seemingly irrational behavior)”, *Applied and Computational Mathematics*, 2014, Vol. 13, No. 3, pp. 275–298.

- [8] J. Lorkowski and V. Kreinovich, “Granularity helps explain seemingly irrational features of human decision making”, In: W. Pedrycz and S.-M. Chen (eds.), *Granular Computing and Decision-Making: Interactive and Iterative Approaches*, Springer Verlag, Cham, Switzerland, 2015, pp. 1–31.
- [9] J. Lorkowski and V. Kreinovich, “Fuzzy logic ideas can help in explaining Kahneman and Tversky’s empirical decision weights”, In: L. Zadeh et al. (Eds.), *Recent Developments and New Direction in Soft-Computing Foundations and Applications*, Springer Verlag, to appear.
- [10] E. Rodriguez and G. Chang, “Expected academic performance in a lower level undergraduate structural course”, In: A. İşman and A. Eskiculami (eds.), *Proceedings of the Joint 2017 International Conference on New Horizons in Education (INTE), International Distance Education Conference (IDEC), and International Trends and Issues in Communication & Media Conference (ITICAM)*, Berlin, Germany, July 17–19, 2017, pp. 1323–1331.
- [11] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.