

A Detailed Weighted Analysis Approach Towards Course Counseling System For Education

Anush S. Kumar*, Sreedhar Radhakirshnan[†] and Sushrith Srinivas Arkal[‡]

Department of Computer Science and Engineering

PES University, Bangalore

Karnataka, India 560085

Email: *anushkumar27@gmail.com, [†]sreedhar1895@gmail.com, [‡]sushrith.arkal@gmail.com

Abstract—In this project we introduce a text analytics based approach to build a course counseling system for education. While already existing approach of a predictive linear numerical model provides some level of accuracy to estimate a student's inclination towards a course before its commencement, there are limitations such as non consideration of impact of prerequisites of a course in the model. Even standard consideration of prerequisites does not provide detailed weighted analysis as prerequisite is a boolean variable. In this project we quantify the impact each prerequisite has on a course and use this measure to provide a more detailed predictive model for course performance estimation of each student. This model not only helps a student plan his/her approach in a course in a better manner, it also helps professors understand each student's inclination in the course before course commencement. This further has an impact on course planning and can potentially improve education management. We aim to make the model dynamic such that the model alters based on student performance in test, assignment and project in the courses opted by the student. Future work includes using the model to generate a course approach plan for each student as well as a directed project allocation based on student interest given a repository of projects by the professor.

I. INTRODUCTION

For a course, there may exist dependencies (often represented as prerequisites) that give us an idea of the skills and knowledge required to perform in a satisfactory manner in the particular course. Thus, it stands to reason that using these dependencies, there must be a way to predict a student's performance in the course (of course prior to having attempted the course) using the dependencies and an data of how well the student has performed in the dependencies of the course. For example, a course on algorithms depends on a solid foundation in discrete mathematics and logic, data structures, etc. If it is ascertained that the student has strong skills in the dependencies, then their chances of performing well in said course increases. Thus, it also stands to reason that the student need not spend as much time studying that subject to be able to perform well. In this manner, we can generate the most optimal study schedule for a student to be able to perform well academically.

While it may be the case that the above process seems simple enough for a person to do themselves, often times, in a classroom setting, it is difficult for a single teacher to accurately judge every single one their students' aptitude and determine the appropriate course of action to improve their

performance. In addition, this process would require that the teacher get to know each student's aptitude and skills before they can advise the student on fruitful application of their time and efforts. Thus, there would be considerable benefit to having a system by which we can advice a student on the necessary actions that need to be taken to ensure that their performance is satisfactory, prior to have any data on their current performance in the course. Such advice would come through the form of a 'study schedule' that instructs the student of the time needed to be spent in order to achieve academic success.

Our problem is to gain an understanding of a student's aptitude and proficiency in various courses through various data sources like test scores, project grades, etc. and determining the dependencies between a course and its prerequisites, we use this information to generate a ordered list of courses, ordered with the most favorable course first, the course that would be better suited to for that particular student. The first step would be to determine the appropriate metric to measure the level of dependencies between the course and prerequisites, followed by determining the model needed to solve the problem with the dependency measure and student's previous performance on the prerequisites. Time permitting, we may also be able to increase the granularity the of the dependency analysis from course level to course unit/chapter levels and to use the data to more accurately determine the student's study schedule.

II. RELATED WORK

In this section we will look at some of the works that has taken place pertaining to certain parts of the problem that we are trying to solve, but course wise analysis(considering more insights on the course) is not yet explored in the research literature. . After doing some considerable literature survey on IEEEExplore[1] and other sources[2] we seen many approaches, predictive models and recommender systems to predict student performance before commencement of the academic year. [3] [4] [5] With the existing approaches we can only predict the performance based on previously obtained grades of the student, which might not be a sufficient feature that would predict the outcome of that particular course performance. Particularity this paper[5] interests as this give in insight is to how data mining can help predict the effect of

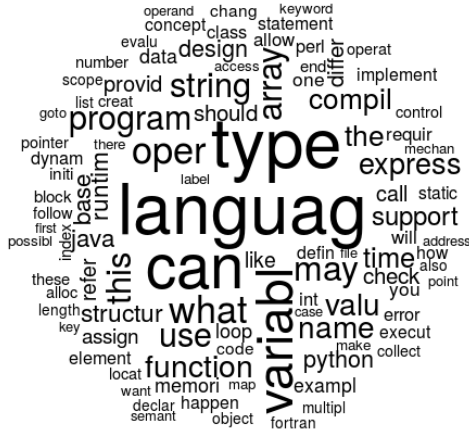


Fig. 1. Wordcloud done on notes of course - **Principles of Programming Languages**

difficulty of a particular course on student's performance, the take away from this study is that it is really important to choose the right course of liking to avoid unnecessary pressures to score.

III. DATA SAMPLE

The data source we plan to use on this project is university provided record's of the students' scores and/or grades in various computer science courses (including courses on data structures, algorithms, database management systems, etc.) along with course material for each course, which will be extracted from NPTEL's online video course transcripts, the university's course information records, and from the appendix and index of reference textbooks. We also plan on extending our system to process information from various other sources such as professor ratings, student course difficulty ratings, etc.

The data on students' grades/scores will not be compared against each other in our project, but will be used to assess the 'history' of each student, to determine the predicted performance of the student in a particular course. The course materials will be used to determine the level of dependency the course may have on it's prerequisites. Other data sources such as professor ratings will be used to determine the difficulty level the student may expect in a course as taught by the professor.

IV. INITIAL ANALYSIS

So far, we have done some initial analysis on the course material for various courses, including a course on computer networks, data structures and algorithms, programming languages.

The first step we took was identifying viable sources of course material, where in we could get enough text



Fig. 2. Wordcloud done on notes of course - **Data Structure and Algorithms**

and material that describes the content and possibly indicates the dependencies of the course. So far, we have decided on using National Program for Technology Enhanced Learning(NPTEL)[6] online video computer science course transcripts to as they are actual transcripts of lectures taught to students, and as such should have all the content relevant to the course, and will reveal the extent to which a course is dependent on it's prerequisites.

As an initial first step in analysing the course material, we have generated wordclouds using R[7] to gain an idea of the types of common words (are the words related to a prerequisite or are they unique to the course), and the frequencies. From these wordclouds, we can see that the words unique to the course, in the courses we have analysed, seem to dominate the words related to any of their prerequisites. In this case, this seems to imply that the dependency levels seems to be low-mid range, compared to the entire content of the course material.

V. PROPOSED SOLUTION

We look to address the issue of individual student interest and aptitude understanding before and during course commencement when student professor ratio is many is to one. This system also assists Professors to understand his class at an individualistic level from an aptitude and interest point of view.

Most approaches in this domain deal with performance estimation purely on the basis grades obtained. Our approach is a hybrid of both text analysis as well as numerical analysis which has not been done before as to our knowledge. The text analysis is used to quantify the percentage/level of each prerequisite for a course and the results of the same improves the accuracy of the numerical predictive model. Analysing (Figure3 depicts) the dependency between the two gives us

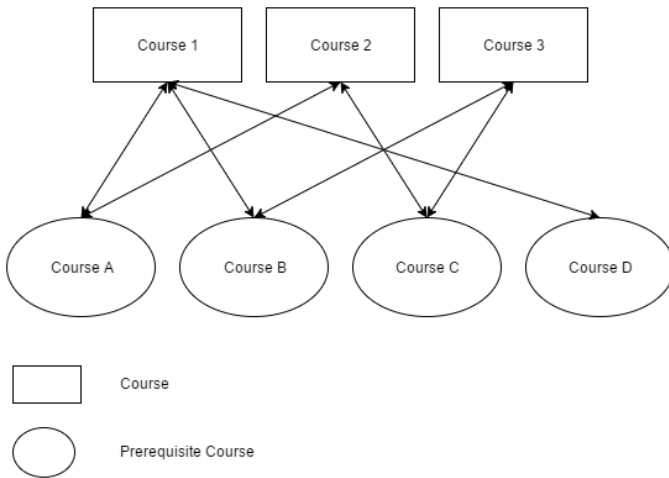


Fig. 3. Course Dependencies

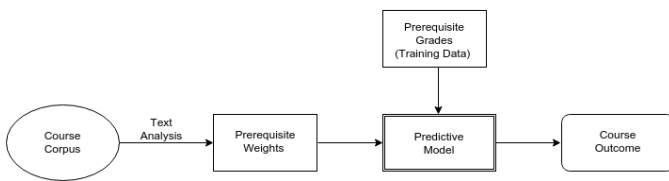


Fig. 4. Proposed System Workflow

greater insights is to the extent of which the prerequisite course actually affects the new course, which gives us an extra feature to enhance our model's performance. Also while most predictive systems in education deal with just prediction, our system is dynamic in the sense the model varies as and when test, project and assignment scores are updated thereby acting as a counseling system.

The text analysis is primarily based on word frequency generation and entity matching for each course across other courses. Based on the frequency and percentage of match, a metric is developed to determine the weight of each prerequisite required. What also makes this approach unique is the fact that we are not declaring the prerequisites before the analysis and determination of prerequisite weight takes place in an unsupervised manner. Upon generation of the weights post the text analysis we look to build a predictive model using these weights to provide a more detailed estimate of a student's current inclination of the course.

VI. CONCLUSION

In this project we used a hybrid text and numeric analysis approach to build a course counseling system for education. In the process of doing so we introduced an unsupervised approach to quantify percentage of prerequisite required for a course. The developed model does not have a single task based focus and caters to both student as well as professor needs. Future work includes project matching based on student interests given a repository of projects. This add-on will thus provide a complete course counseling system for education

particularly in the undergraduate and graduate level. Keeping the model as the brain at the back-end we can develop a cross platform application (both web and mobile) for a realtime analysis and usage of the solution, with this implementation we are exposed to a whole new set and type of data which can be used to further enhance the model's capabilities. We will have the data of popularity of each course, trends in selection of the course and much more data.

ACKNOWLEDGMENT

We thank our professors from PES University who provided insight and expertise that greatly assisted the research, although they may not agree with all of the conclusions of this literature survey. We would like to show our gratitude to the Dr. Viraj Kumar, Professor, PES University, Dr. Gowri Srinivasa, Professor, PESIT South Campus and Prof. Shreekanth M. Prabhu, Professor, PES University for giving us this opportunity and sharing their pearls of wisdom with us.

REFERENCES

- [1] IEEEExplore. [Online]. Available: <http://ieeexplore.ieee.org/>
- [2] Google Scholar. [Online]. Available: <http://scholar.google.com/>
- [3] K. Kongsakun, "An improved recommendation model using linear regression and clustering for a private university in thailand," in *2013 International Conference on Machine Learning and Cybernetics*, vol. 04, July 2013, pp. 1625–1630.
- [4] T. Mishra, D. Kumar, and S. Gupta, "Mining students' data for prediction performance," in *2014 Fourth International Conference on Advanced Computing Communication Technologies*, Feb 2014, pp. 255–262.
- [5] K. Kaur and K. Kaur, "Analyzing the effect of difficulty level of a course on students performance prediction using data mining," in *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, Sept 2015, pp. 756–761.
- [6] National Program for Technology Enhanced Learning(nptel). [Online]. Available: <http://nptel.ac.in/>
- [7] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org/>