International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Review on Recommender Systems for University Admissions

Aditi Bhide[1], Karan Karnik[2], Pradnya Badge[3], Varun Joshi[4], Vina M. Lomte[5]

[1, 2, 3, 4]Department of Computer Engineering, R.M.D. Sinhgad School of Engineering

[5]Assistant Professor & Head, Department of Computer Engineering, R.M.D. Sinhgad School of Engineering

Abstract: ***Data mining is the process of exploring and discovering patterns in large datasets. It involves analyzing a data set from different aspects and summarizing it into usable and useful information which can be used to minimize costs and increase efficiency. Students are unaware of the best institute suited for them due to insufficient information. A large number of students apply to universities for admissions in various streams. This paper explores the various recommender mechanisms that use data mining and neural networks to tackle university admission problems.***

Keywords: data mining; recommender systems; neural networks; university admissions;

1. Introduction

University College education plays an important role in a student's life. Hence, choosing the right college for the right student is a highly challenging task. Although many universities have adapted to centralized admissions, the system is not efficient and a lot of time is spent in securing an admission in the institutes. This is primarily due to large number of applications being submitted to universities every year.

Recommender Systems are data filtering mechanisms [1] which focus on overload of information and filter out relevant chunks of data according to user preferences. Such systems have the ability to predict if a user may be interested in certain items based on their profile [2]. Such systems are widely used in e-commerce environments where the system is required to recommend items based on their shopping history [3].

These models can be applied to university admissions process as well. Ragab, Mashat and Khedra [4]have proposed two recommender systems that work together for high performance. The first system assigns the students a track for a preparatory year. The second system recommends the college based on whether the student has passed the first module or not. The algorithm uses previous historical GPA scores for predicting the probability of admission.

The rest of the paper is structured as follows: Section 2 gives an overview about the related work in using recommender systems for university admissions. Section 3 discusses the models proposed by Ragab, Mashat and Khedra [4] which are based on Data Mining and Knowledge Discovery Rules. It also discusses a model proposed by Simon Fong, YainWhar Si and Robert P. Biuk-Aghai[] which is based on Neural Networks and Decision Tree Classifiers. Section 4 compares these models. Finally, the conclusions are presented in Section 5.

2. Related Work

The system proposed by the authors Simon Fong and Robert P. Biuk-Aghai uses a composite model of neural networks and decision tree classification for university admission recommender system[5]. Every year, thousands of students apply to different universities. The universities have to choose the most appropriate students from all the applicants which can be a very intricate task. Various factors such as academic scores, gender and caste, proportion of local and overseas students etc. are taken into consideration while accepting the applications. In the proposed system the author suggests that various other factors such as academic performance can be contemplated from the secondary school that the students have been to. The students can be broadly classified into three main subgroups – 1) Top students that can be recommended for the reputed universities without any entrance examinations, 2) average students who can apply for various universities by appearing for their entrance exams and 3) low performing students. These subgroups can then be considered at the time of admissions. Thus, using this system, the universities can select the most suitable students for the course. Also, students can take factors such as difficulty, duration, cost etc. of the degree course into consideration while applying to the university [5].

Greg Linden, Brent Smith and Jeremy York have thoroughly described the recommendation system used by Amazon.com [6]. A recommendation algorithm called item-to-item collaborative filtering is used, which is independent of the number of customers or the number of products in the item-catalog. Real-time, high quality recommendations can be made even on massive data sets using this recommendation algorithm. In most other recommendation systems, a set of customers who have purchased and rated the same products as the user are found out and accordingly recommendations are made. However, in item-to-item collaborative filtering, instead of customers, the items that the user has purchased and rated are identified and similar products are recommended [6].

An electronic commerce recommendation system proposed and implemented byShu Liao, Tengyue Zou and Huiyou Chang[7] which uses association rules and sequential rules to recommend the appropriate products. Association rules give information of the products which are often bought together whereas; sequential rules provide information about the sequence in which different associated products are

bought. In most of the existing recommendation systems, only the association rules are considered and validated to make suggestions, however, in this system the authors have used both, association rules and sequential rules simultaneously to get more suitable and appropriate recommendations.

GopinathGanpathy and K. Arunesh have proposed two new recommendation models which are Rank-Reco and Test-Reco based on ranking and testing measures respectively[8]. In the first method, based on ratings provided by the user, a smart system is constructed. These ratings are then converted into rankings. The similar aspects are identified in this method by the numbers of ratings for the items by the users. In the second method, the inadequacy problem is taken into consideration.A large number of items are presented in e-Commerce applications, but sometimes user only selects and rates afew items. This leads to sparse entries in the rating vector. In this case the significant user tastes are predicted based on paired t-test for difference of means. This two models generate recommendations for web user.

A new concept based on the field of web usage mining, a new recommender system for user rating and synonyms based ranking of the websites has been proposed by Anamika Rajput and Sushil Kumar Chaturvedi[9]. This paper focuses on data pre-processing which is the first step of web mining. In this system, when users search for a keyword then that keyword is searched along with related synonyms. Websites are displayed based on its search and user is asked to give rating. By user rating that page is upgraded or degraded in the database. To differentiate and to reduce data retrieved map reduce algorithm has been used. This recommender system is efficient and works faster.

3. Recommender Systems for University Admissions

A. HRSPCA: Hybrid Recommender System for Predicting College Admissions

Admission to a university is a crucial factor in determining the success in carrying through the education. A large number of students are admissible to universities and the criteria for the same includes a lot more than just merit. Factors like student background, gender influence the eligibility of the students for a particular university.Ragab, Mashat, Khedra[4] propose an admission system which makes use of two recommenders to achieve concise and accurate results. The main goal is to provide the students useful courses and guidelines in a way that is best suited for them.The system comprises of two recommenders, for two different mechanisms, one for recommendation and the other for prediction. They are cascaded together and applied along with data mining techniques[10]. The main focus of the former is to provide guidelines for the students at the elementary stages Data mining techniques are applied to find similarities among a huge number of applicant students data[8]. After the completion of the process the students are represented with the courses and colleges appropriate for them while considering their grades and other criteria.

To achieve the desired outcome, the system is divided into stages all of which are interrelated and only those students which successfully pass these stages are represented with the calculated interpretations.

1) Web Portal

The system structure consists of an interactive and correlative web-portal to ensure communication between applicants and the system provides facts and data about the search techniques. The applicant student is provided with a login ID and a password. The student must enter his/her personal data which includes name, date of birth, address, email, mobile number, social status, disability status and details regarding subsidization. Information about the school, the merits and grades of the student are registered. After these preliminary tasks, the available university courses are provided and selection must be done on the basis of the desires in priority. A suitable parameter is obtained from the entered data which can be used for further calculations.

2) Enrollment stage

The particulars are registered in the system successfully after the student enters the data. The user id and password is provided to the student so as to enable access to the system and continue entering the necessary data in the further stages.

3) Documentation and auditing

All the necessary certificates and credentials are to be provided by the students. Data clearing is cultivated in this stage, i.e. verification, validation, conversion and validation of the data takes place. This exercise is to verify the correction authentication of the student's name, age, nationality, etc. A mail is sent to the students if any mismatch in the data is found. After this procedure, the validated data is stored in the system database. The documents of the students which are accepted are then sent further to the track recommender.

4) Track Recommender

The TR is a fundamental track sorting recommender. It consists of a sorter and a filter. Tracks are made according to the study tracks available provided by the universities. The sorter sorts students to these several tracks. After the sorter finishes its task, the filter is used to re-organize the students in two classes; the students who pass the courses successfully are sent to college recommender to be allocated to certain colleges and the students who fail are dropped and are postponed of services till they succeed.

5) College Recommender

The recommender contains two components, namely, the classifier and the allocator. Service is provided to those students who completed the courses triumphantly. The classifier categorizes the applicants according to the qualifications and the other components like the gender, background etc. the students are then allocated the respective universities in a just manner [11]. Pattern discovery rules of data mining are used for classification. For example, these rules can be used to distinguish the resident students from the immigrant students. Here a pseudocode of feature selection and knowledge discovery is used.

6) College Predictor

The predictor is used as a non-compulsory alternative so as to provide optional help to students regarding the tracks and colleges. The current applicant i.e. student data is compared to the data already available in the system, the historic data. Thus the students are provided with the most suitable colleges [12]. The most probable colleges are represented according to his/her merit and the qualified exam marks in prep year.

This information provides guidelines to the student to arrange his desires in priority.

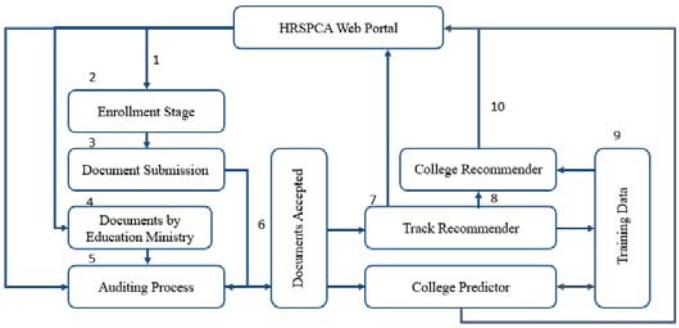

Figure 1: The HRSPCA Architecture

To ensure the smooth functioning of the stages, certain rules are applied to the TR, CR and certain operations are used by CP.

7) Rules for TR

The TR focuses on the prep year criteria and the output allows the students to be clustered in the following five tracks:

R1: Demographic student information:
The data fields documented in the student personal data are governed.

R2: Scientific Track:
Different portions of the qualifying exams are considered- (50% of high school degree) + (30% of aptitude) + (20% of achievement degree)

R3: Literary Track:
(60% of high school degree) + (40% of aptitude)

R4: Affiliated Track
(50% of high school degree) + (30% of aptitude) + (20% of achievement degree)

R5: Colleges with no preparatory years:
The colleges which do not have tracks for the prep years are also considered.

Here, the student data is preprocessed and filtered after the application of four main rules with the tracks. Furthermore, gender classification is applied on each rule.

8) Rules for CR:

Data is obtained from the TR and the CR mainly governs college allocation based on the following criteria:
R1: Passing all prep year courses

R2: Minimum score required for the prerequisite courses offered by the college

R3: Capacity of the college
Thus various parameters can be established for different courses, and depending upon the restrictions, the allocation can be done appropriately.

9) CP Operations:

The qualified students have a high probability of admission whereas the admission of some of them may depend on other factors. For example, more students can be admitted if the college capacity increases. The rest may have no chances for attending colleges.

Data mining is used in almost all of the stages; associations and sequences or correlations are used[10]. The data is therefore represented in a close-packed manner to improve efficiency. Data clustering and classification are used to categorize the data present into a format which is domain-specific[13]. The recommendations are then obtained as the final output.

B. Hybrid model of Neural Networks and Decision Tree Classifiers

Fong, Si &Biuk-Aghai [14] believe that university admissions are dependent upon much more rather than just test scores. Admissions also rely on student backgrounds and changing finances and academic interest. Existing systems depend largely upon yes or no decisions and linear programming [15].

Student profiles are built by modelling attributes of the student to the weights of the neural networks. Historical data is used as a training data set. These networks are adaptive; they affect the categorization based on change in weights as and when the system goes through additional training on recent data [16].

Student data was observed during counselling sessions to determine the basis on which students choose a university. Inputs to neural network are decided from these observed criteria. Each group of students applying to universities has a distinct neural network system. Hence a huge amount of computations is required and training the network takes up a lot of time.

A hybrid model is proposed which takes in a large number of input records *m*, and outputs results into *n* groups. For admission to universities, the system uses *m* student profiles which contain demographic and academic information and *n* predicted universities into which a student may get admission into and the procedure through which the student should apply.

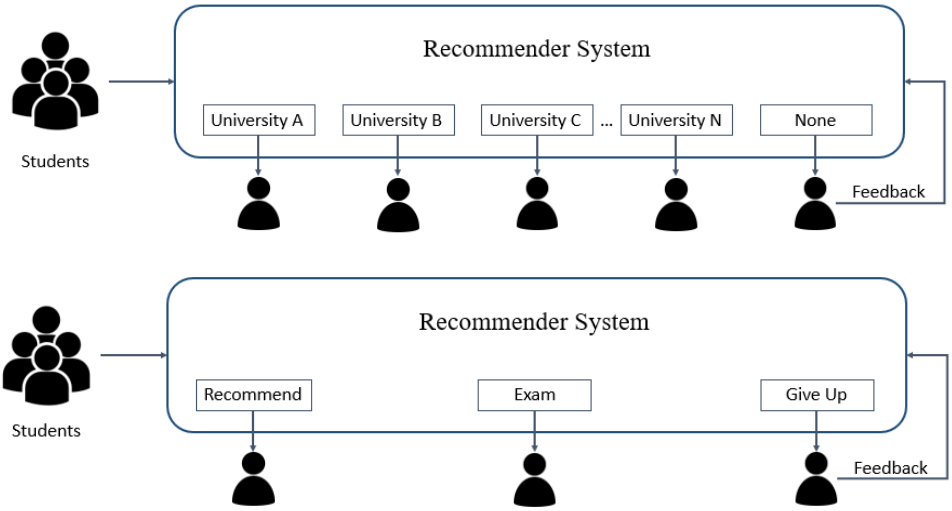

Figure 2: Classifying Students into different channels

Step1): Back-propagation algorithm [17] is applied to train and construct a learning model and calculate the importance of input in correspondence to the output, such as university recommendation. Sort the input into a list as per feature.

Step2): Input variable is added in the list produced by neural network to check and validate performance of C4.5 algorithm [18] and error rate is noted. If theerror rate of the input at iteration *i*+1 is found to be higher than that of *i*, stop. Continue till full list is checked.

Step3): The rules generated by the C4.5algorithm are checked by measuring its performance. When it satisfies the criteria set by the user, it is ready to be used with new data.

Selection of correct variables most relevant to admission into a university is an important factor to achieve high accuracy of prediction.

In this hybrid recommender model aback-propagation neural network algorithm was used to sort out the relative important input variables from the available ones. Once the neural network identifies the important input variables and eliminates the rest, the chosen variables are exported to the C4.5algorithms for generating a decision tree. A classification tree is then built by the generated rules of university admissions.

We can find the most favorable values of the weights that minimize the error between the measured and the output performance parameters from the neural network model. A relative importance of input concepts is used to establish a measure of significance for each input variable by defining the range of the values between 0 and 1 so that higher values are associated with more important variables after modeling in a neural network algorithm.

From the neural network model, we can find the optimum values of the weights that minimize the error between the measured and the evaluated (output) performance parameters. After modeling in a neural network algorithm, a relative importance of input concepts is used to establish a measure of significance for each input variable by defining the range of the values between 0 and 1 so that higher values are associated with more important variables.

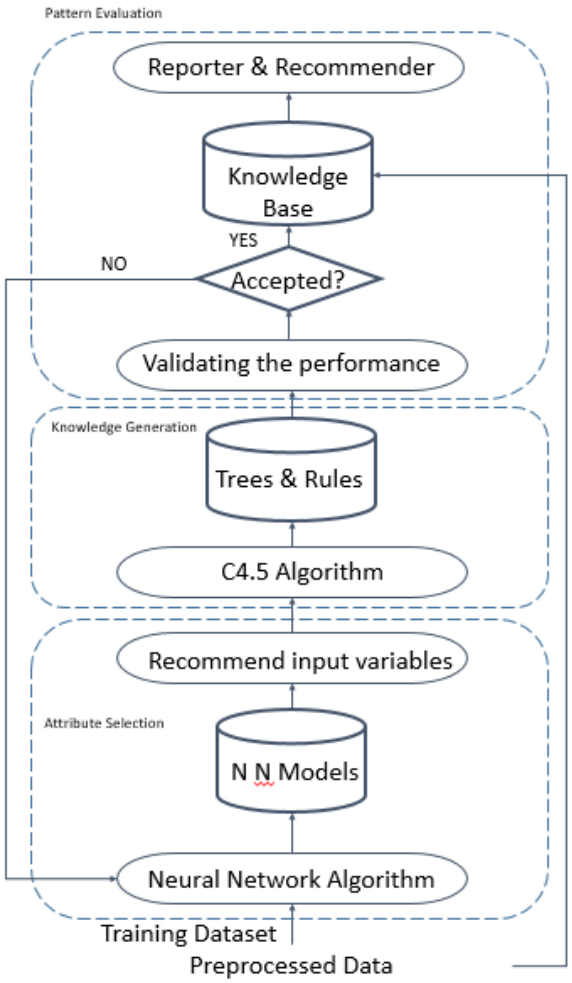

Figure 3: Workflow of RSAU

The loop procedure chooses best inputs for categorization.
1. i = 1;
2. n = number of attributes,
3. $attribute_i$ represents the rank *i* attribute in the list generated by neural networks algorithm
4. Repeat

4.1 select attribute$_1$ to attribute$_i$ as input attributes to build classifier by C4.5 algorithm
4.2 calculate error rate ->e_i
4.3i++
5. Until i = n or e_i> $e_{(i-1)}$

Output: The optimization classifier model with *i* attributes.

4. Comparision

The HRSPCA system was implemented using live data and the web portal was accessed by 66,000 students during the year 2012 by Khedra et. al. The students who passed the tests successfully and were allocated standard colleges and universities were close to 16,000. The students who were rejected may appear again. The practical experimental results are shown as follows:

a) TR Outputs:

The student marks are indicated in the output which is calculated in the sorting process according to the weightage computed. The weights are calculated by taking the product of hours and degree. The students are classified using tracks chosen and the gender is also taken into consideration.

Out of a total of about 19,000 students 28% were recommended to enter literary track, which included both male and female students while 58% students for scientific track. The scientific track included about 2,900 females and 7,000 males. 18% were recommended affiliated track and only 600 students were recommended no prep year.

b) CR outputs

The capacity of the colleges is a major factor which is considered while students are to be recommended for all available university colleges. The outputs include classification, again, according to gender. The output includes recommended student's capacity and the colleges than can be acquired.

The recommended outputs are compared to the colleges actually allocated to the students. The system trustability is achieved and ensured when there is more positive response from the students.

Neural Networks are being used extensively in the industry for many years now with adequate library support and implementations [19]. They are easy to conceptualize. The performance of the above hybrid algorithm of neural networks and decision tree classifier was tested with live student data.

One of the advantages of a decision tree classifier is that it takes a significantly less amount of time to learn a large dataset on a medium spec workstation. The proposed classifier was found to outperform other approaches such as the C4.5 Classifier and the neural network classifier. Error rates were found to be significantly higher when the number of instances was less than 1000, implying that a certain amount of instances are required to positively affect performance.

Neural networks using the back-propagation algorithm took about 20 minutes on the above mentioned platform. The hybrid model proposed by Fong, Si, Biuk-Aghai [14] uses back-propagation algorithm for the feature selection and then applies C4.5 algorithms to generate rules for university recommendation.

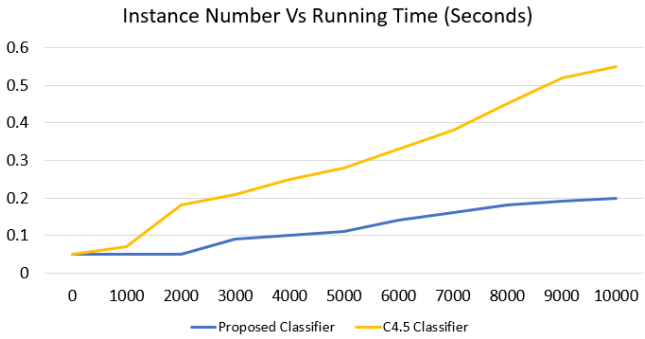

Figure 4: Instance number vs running time comparison of C4.5 Classifier

Although Neural Networks have a proven track record of success for certain problem domains, but there are simpler and more cost effective alternatives available. Training outcome can be non-deterministic and depend crucially on the choice of initial parameters. Training a neural work can take up a large amount of time depending upon the size of a dataset.

Although the real students' data used was from Macau, the design of the recommender is generic and applicable to educational systems in other countries [20].

5. Conclusion

University admissions are an important part of a student's life. Hence, a service must be available that guides the student properly. The system must be fast, accurate and light-weight to achieve higher efficiency. The system must be able to recommend a university based on the student's merit and choice. The HRSPCA system allocates students into suitable streams as well as suitable colleges. The HRSPCA works on cascaded hybrid recommender based on knowledge discovery and data mining which leads to accurate prediction and recommendation.

Neural networks and decision tree classifiers were used to develop a Recommender System of Admission to University(RSAU). It analyzes student academic history to calculate chances of admission to a certain university. The experiments showed that a hybrid decision tree and neural network approach improves accuracy in admission to university classification task and performs substantially better than a single decision tree orneural network.

Although these systems were tested on Saudi Arabian and Macau Universities, they are generic and can be applied to any other universities also.

References

[1] Joseph A. Konstan and John Riedl "Recommender systems: from algorithms to user experience," Springer Science+Business Media B.V. 2012, DOI 10.1007/s11257-011-9112-x.

[2] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommnedation systems: Principles, methods and evaluation", Egyptian Informatics Journal (2015) 16, 261–273.
[3] J. Ben Schafer, Joseph Konstan, John Riedl, "Recommender Systems in E-Commerce," GroupLens Research Project.
[4] Abdul Ragab, Abdul Mashat and Ahmed Khedra, "HRSPCA:Hybrid Recommender System for Predicting college Admission",2012 12th International Conference on Intelligent Systems Design and Applications (ISDA) 2012,107- 113, DOI: 10.1109/ISDA.2012.6416521
[5] Simon Fong, Robert Biuk-Aghai, "An Automated University Admission Recommender System for Secondary School Students", The 6th International Conference on Information Technology and Applications (ICITA 2009)
[6] Greg Linden, Brent Smith and Jeremy York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering", Industry Report 2003.
[7] Shu Liao, Tengyue Zou and Huiyou Chang , "An Association Rules and Sequential Rules Based Recommendation System", DOI: 10.1109 / WiCom. 2008 . 2592 , Conference: Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08.
[8] Anamika Rajput and Sushil Kumar Chaturvedi, "User Rating Synonyms Based Modified Ranking Technique for Recommender System", International Conference on Computational Intelligence and Communication Networks, 2015
[9] Gopinath Ganapathy and K.Arunessh, "Models for Recommender Systems in Web Usage Mining Based on User Ratings", World Congress on Engineering VOL I, July 6 - 8, 2011, London, U.K.
[10] J. W Seifert, "Data Mining: An overview" CRS Report for Congress, December 16, 2004. http://www.fas.org/irp/crs/RL31798.pdf
[11] X. Su and T.M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques" Advances in Artificial Intelligence Volune 2009 Article ID: 421425
[12] A. H. M. Ragab, A. F. S Mashat, A. M. Khedra "Artificial Aided College Recommmender" Paper under processing, KAU, DSR, Technical Funded Project No. 20-611-D1432, unpublished
[13] D. Mican and N. Tomai "Association Rules based Recommender System for Personalization in adaptive Web based Applications." http://gplsi.dlsi.ua.es/congresos/qwe10/fitxers/QWE10_Mican.pdf
[14] Simon Fong, Yain Whar-Si, Robert P. Biuk-Aghai, "Applying a Hybrid Model of Neural Network and Decision Tree Classifier for Predicting University Admission"DOI: 10.1109/ICICS.2009.5397665 Conference: Information, Communications and Signal Processing, 2009. ICICS 2009.
[15] Gottlieb, E. 2001, "Using integer programming to guide college admissions decisions: a preliminary report", *Journal of Computing Sciences in Colleges*, Volume 17 , Issue 2, Pages: 271 – 279
[16] Wasserman, P. D. 1989, "Neural Computing", *Van Nostrand Reinhold*, New York.
[17] AP Engelbrecht, I Cloete and JM Zurada. 1995,"Determining the significance of input parameters using sensitivity analysis. In Proceedings of the International workshop on Artificial Neural Networks: From Natural To Artificial Neural Computation (June 07 - 09, 1995)", *Lecture Notes In Computer Science*, Vol. 930. Springer-Verlag, London, Pages: 382-388
[18] Harvinder Chauhan, Anu Chauhan "Implementation of decision tree algorithm c4.5" International Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013 www.ijsrp.org/research-paper-1013/ijsrp-p2278.pdf
[19] Neural Networks pros and cons https://www.cs.cmu.edu/~schneide/tut5/node24.html
[20] Alexander, S., Clark, M., Loose, K., Amillo, J., Daniels, M., Boyle, R., Laxer, C., & Shinners- Kennedy, D. 2003, "Case studies in admissions to and early performance in computer science degrees", In ITiCSEWGR' 03: Working group reports from *ITiCSE on Innovation and technology in computer science education*, Pages:137-147, ACM Press