An Empirical Analysis of Skewed Temporal Data for Distribution-based Course Similarity

Tao Xie
Faculty of Education
Southwest University
Chongqing, China
xietao@swu.edu.cn

Chaohua Gong
Faculty of Education
Southwest University
Chongqing, China
chaohua@swu.edu.cn

Geping Liu
Faculty of Education
Southwest University
Chongqing, China
gepingl@swu.edu.cn

ABSTRACT

Time series data that exhibit skewed distribution is a common and important issue related to advanced model adoption, which however may be mis-specified when the data become extremely large and completely stochastic. This study adopted an experience-to-model approach in order to address the data skewness problem in educational data mining and in parallel explain the practical pedagogical meaning of data distribution patterns. To do this, we first specified a proper analysis granularity with respect to temporal data and provided evidence of its non-normality, and finally handled the skewness by correlating it to gaussian mixture models. We performed a scalable model by adaptively selecting the parameters and discussed the similarity measure based on probability density distribution.

Keywords

Data skewness, temporal pattern, data transformation, e-learning

1. INTRODUCTION

In recent years, big data in education is becoming a new driving force and playing an increasingly important role in educational research and practice[1]. The mining of big data in education is beneficial for educators and organizations to understand the learning patterns of students, optimize curriculum design, gain insight into student characteristics, provide high-quality educational decisions, and finally improve students' academic standards[2-4].

One of main challenges, however, is that the data is not always of normal distribution, which makes many standard approaches limited and corresponding results not robust. Many studies either ignored the existence of this challenge or simplified the assumptions of research conditions. Pearson correlation, for example, is used for testing linear dependence between a couple of variables assuming the data is small and has a normal shape. But the model can be easily mis-specified because the feasibility of this assumption is weakened when the data become extremely large and completely stochastic. Besides, many scholars use machine learning algorithms to classify the time series data without paying much attentions to the data distributions, leading to seriously inaccurate results due to the fact that the performances of classifiers are subject to the data presented to it during training session and many attributes of the data are not balanced[5]. The imbalanced data distribution is usually described by a skewness coefficient in statistics representing an asymmetry from the mean of a data distribution. Time series data that exhibit skewed behavior is a common and important issue related to advanced model adoption of educational data mining[6].

This study adopts an experience-to-model approach in order to address the data skewness problem in educational data mining and in parallel explain the practical meaning of data distribution patterns. To do this, we first summarized coarse-grained observations on temporal data that collected from online courses, and then discussed the non-normality of time series data in education based on carefully selected granularity. Finally, we provided a tutorial to handle the skewness by correlating it to gaussian mixture models.

2. RELATED WORK

Temporal data has been studied in many research domains, while there is only a little literature has been documented in the educational domain. The current research on educational temporal data has mainly focused on online courses and metacognition. For example, authors in[7] proposed a temporal modeling approach for students' dropout prediction in MOOCs, authors in[8] mined temporal characteristics of learning behaviors from e-learning systems, and authors in[9] obtained sequential and temporal characteristics of self and socially regulated learning. Many of these research omitted the distribution assumption of data samples, which may lead to improper explanations related to statistical values. As is known, the Gaussian distribution is well known and widely applied by assuming that the aggregate effect of many individual independent components tends to be distributed with symmetrical bell curve. However, the use of Gaussian-based statistics can result in substantial error if problems are involved a lot of skewed data[10]. The assumption of homogeneity of variance indicates that the variance of the variable remains constant over the observed range, which may not be the truth in most research scenarios. Although the current statistical software packages provide tools to test the normality assumptions, and a lot of literature have documented to use multiple regression model and ANOVA model for modest violations to these assumptions[11]. A more effective way, however, is to transform data to improve normality of independent variables when substantial non-normality is present. Data transformations can improve normality of a distribution and equalizing variance in quantitative analysis of data. For this reason, this study will conduct experiments based on this approach. In previous works, the transformation approaches include adding constants, square root transformation, log transformation, scales, inverse transformation, arcsine

Tao Xie, Chaohua Gong and Geping Liu "An Empirical Analysis of Skewed Temporal Data for Distribution-based Course Similarity" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 710 - 714

transformation and Box-Cox transformation[12]. Many of these approaches are showing good properties of distribution symmetry.

Besides, the data skewness problem has been extensively studied in communities of time series mining. It is claimed that the skewness has significantly influence on the performance of algorithmic tasks[13], where the authors have to detect the degree of skewness to determine the characteristics of the dataset distribution. To explore such influences, the relationship between data skewness and accuracy of data mining models has been examined in[14] and[15]. Because the irregular sampling of data sets is often encountered in time series, the measures of skewness are also of great interest and have become especially important while conducting data cleaning and data preprocessing. Therefore, authors in[16] devised an approach to transform the time series segments to produce new ones so that the new ones can be analyzed using standard methods, which is in essence consistent with data transformation techniques as stated above. This study uses quadratic square root approach on temporal data and compares the difference in the developed metrics between original and transformed data on online courses.

3. DATA

3.1 Data Collection

There are 14 million video-viewing logs collected from 57,717 students. We choose the top 7 video courses with the most in-course interactions, which are respectively Introduction to Mao Thought (MS), Political Economy (PE), Linear Algebra (LA), Enterprise Financial Management (EF), Marketing (MM), Microcomputer Principle and Interface (MI), and Health Assessment (HA). Finally, we keep information of 7,341 students. One-way ANOVA shows that the differences between groups in terms of the continuous variables are statistically significant at the 0.05 level ($df = 6$, $p=0.00$) and the differences between groups in terms of the age and video-viewing time have statistical significance at the 0.05 level ($df=4$, $p=0.00$).

3.2 Granularity of Analysis

In information systems, time is mainly represented by time points and time intervals. Time describes the moment at which learning behaviors occur, while time interval describes the length during which the behaviors last. They are used to present a certain chronological order, cycle characteristics, and time association rules. Since the current analysis unit is the temporal data, time-related information of interest is abstracted. Each student has plenty of but usually intermittent interactions with systems. In order to summarize the statistic distributions, this study focuses on the total time during which a player is always in the playing state. The video-viewing time is an absolute measure representing the length of content students learn, while we also consider a relative measure called the viewing completion ratio, which is the proportion of video-viewing time with respect to the total video length and represents the progress of content consumption.

3.3 Preliminary observations

The most active period for students watching videos is from the November of the second half of the year to the early January of the following year. The effective learning days of the week are workdays, and ineffective learning days are weekends. The study period of the day is mainly from 9 am to 6 pm. During mealtime and other breaks such as the evenings, students rarely watch videos. Students who use mobile devices have different temporal patterns.

For each course, the cumulative playing time of most students is less than 1 minute. This shows that this part of students is not advanced students[17], Their behavioral pattern can be attributed to “zapping style” according to[18]. There are also some students whose cumulative playing time exceeds the total length of the video and the corresponding learning completion rate is greater than 1, which indicates that these students have played the complete video from the beginning to the end or watched specific segments of the video repeatedly. In other words, their watching pattern can be attributed to repetitive style[18].

As the length of time increases, the probability density first decreases rapidly to a specific value, then reaches a peak at a faster rate and produce a thick tail. For each course, the peak of the probability density curve is relatively close in time and has a similar co-increasing or co-decreasing trend. This shows that the students' learning of the courses shares a similar distribution pattern.

The average of viewing time and the viewing completion rate are both close and low. The average cumulative viewing time for each course is about 13 minutes, and the average completion rate of video viewing is about 40%. These two values reflect the phenomena reported by most MOOC studies: high dropout rates and low resource utilization. It also shows that, compared to non-educational videos, educational videos have specific non-linear viewing patterns and a clear cognitive search intent[19].

4. RESULTS

4.1 Skewness

We perform a Kolmogorov-Smirnov normality test on the selected 7 courses. The established null hypothesis is that the viewing time or viewing completion rate conforms to a distribution of a specific normal shape. At 95% confidence and 0.05 significance level, we calculated two-tailed p-values for the two indicators which are showing equivalence to 0.000. In addition, we calculate the D statistic, which tells the maximum distance of the cumulative distribution function between the data distribution and the fitted normal distribution. More intuitively, it quantifies the magnitude of the difference between two distributions. For each course sample, the D value is large.

Besides, we observed the kurtosis and skewness coefficients. We can find that the probability density curve has a sharp peak and right-skewed shape. The right skewed distribution has a property that the mean value in the horizontal direction is greater than the median and mode[20], and the absolute value of most skewness is greater than 1.96 times its standard error, which indicates that the skewed distribution and the symmetrical distribution have statistics significance. Like the average viewing time, the distribution of viewing completion rates is all right-biased except MM, and the relationship between the median and the mean satisfies the corresponding skewed properties.

Because of data skewness, it is not appropriate to use standard methods that are based on normal distribution assumption. There are generally two methods for processing skewed data. The first method based on fitting a series of models has been implemented in[21], and the current study will try the second approach to obtain more rich features through data transformation.

4.2 Data Transformation

Intuitively, the right-biased distribution of the data causes the probability density of the long tail to change relatively slowly. This means that once students exceed the average viewing time threshold, they will have a higher proportion to invest more time in courses. Conversely, if students' learning time does not reach this threshold, they will have a higher percentage of withdrawal from the course. Indeed, it is observable that many students fall into the second category. In order to make the curve more symmetrical, we compress the spacing of the data. This requires that the spacing of the long-tail portion is compressed faster, while the short-tail portion is compressed more slowly. After trying a lot of models empirically, we found that quadratic square root of the original data can make the data distribution basically symmetrical, and its effect is better than other available methods, such as natural logarithm. We further use the local quadratic regression on the transformed data to smooth the data. Let the range of the temporal variable *X* be *D*. For each sample $x_0 \in D$, we choose a neighborhood of x_0. Fit the dependent variable corresponding to the observations of the temporal variables that fall within this neighborhood using the weighted least squares method. The value of the curve at x_0 is an estimate of the regression function. The results are showing in Figure 1.

large number of samples and thus shows more characteristics of normal distribution according to the central limit theorem[22]. Additionally, EF, MI, and HA courses have smaller AIC values, indicating that they are suitable for a binormal distribution. In order to quantitatively evaluate the improved effect size, we focus on the AIC indicator considering that it imposes more stringent penalties on the complexity of the model compared to other indicators, so that the model we choose not only has the minimum parameters but also prevents overfitting[23, 24]. The effect size can be calculated as the improvement ratio of the AIC value while using the binormal distribution versus the single normal distribution. Results show that the most improved course when using binormal distribution fitting is HA, followed by MI, and the least improvement is MS.

The statistical characteristics of the bimodal distribution indicate that there are different dropout and retention patterns for students in all courses; courses with greater differences in the goodness of fit between single and binormal distributions, say LA and MM, show higher retention rates; instead, courses with smaller differences in goodness of fit between single and binormal

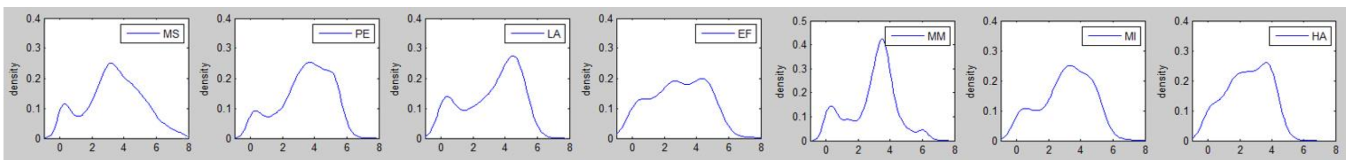

Figure 1. Curve smoothing with respect to video viewing time

distributions such as MS show a more prominent dropout pattern, which should be given sufficient attentions by course organizers and educators.

The transformed time series does not satisfy any single type of distribution but presents the characteristics of a multimodal distribution. In addition to the main peak in the middle of the curve, at least one small peak may appear near the beginning or the end of the curve. In order to evaluate the goodness of fit, we use maximum likelihood estimation to fit the set of curves. Note that the current curve has a more pronounced symmetrical property near each peak than the original curve, which inspired us to try a gaussian mixture model. Suppose a curve approximates a k, $k \geq 2$ normal mixed distribution, the probability density of the curve can be expressed as:

$$p(x) = \sum_{i=1}^{k} \lambda_i p_i(x) \, , \tag{1}$$

where λ_i is the weight of the $i'th$ gaussian distribution, satisfying $\sum_{i=1}^{k} \lambda_i = 1$, and $p_i(x)$ is probability density function of the $i'th$ gaussian distribution, satisfying $p_i(x) \sim (u_i, \sigma_i^2)$. To prevent overfitting due to empirical selection, we consider only the simplest $k = 2$ here, and compare its goodness of fit with a single normal distribution to choose the right fitting model.

Evaluation metrics include root mean square error (RMSE), adjusted R^2, and Akaike's Information Criterion (AIC). They are showing in Table 1 that for all courses the binormal distribution is better than the single normal distribution. Numerically, it always has $RMSE_{k=2} < RMSE_{k=1}$, $R^2_{k=2} > R^2_{k=1}$, and $AIC_{k=2} < AIC_{k=1}$. We can also find that in the two models, the biggest improvement is LA, which means that its binormal distribution feature is more significant. While for MS with little improvement, both single and binormal distributions can be used for fitting. This may depend on the course setting. MS is a campus-wide elective course, which has

Table 1. Evaluation of distributions.

	Single normal distribution			**Binormal distribution**		
	RMSE	Adj.R^2	AIC	RMSE	Adj.R^2	AIC
MS	2.297	0.912	223.8	2.236	0.917	172.9
PE	2.973	0.890	223.9	1.629	0.967	109.6
LA	4.948	0.652	325.8	1.705	0.959	118.7
EF	2.142	0.922	158.3	1.097	0.980	30.5
MM	5.461	0.780	345.6	2.233	0.963	172.7
MI	2.542	0.916	192.6	1.029	0.986	17.7
HA	2.775	0.917	208.7	0.92	0.991	-4.6

Repeating the above experimental process, we find that the distribution of the viewing completion rate is more complicated than that of the viewing time. If we use k mixed distribution for fitting, usually the goodness of fitting can be obtained when $k > 3$. Since the discussion of $k > 3$ is too complicated, we will deal with it by generalizing the model for arbitrary k values in Section 5. In order to reflect more details of the student learning process, we borrow the concept of temporal structure. We argue that it reflects the change of students' time investment when watching the courses, which is helpful for further analysis of students' preferences for in-course parts. It should be noted that viewing time is different from time investment. The former is a static quantity that measures how much time is invested, and the latter is a dynamic quantity that measures the difference in structure of time investment.

4.3 Variation of Temporal Structure

In order to quantitatively describe the difference in the temporal structure of the viewing sequences, we use the coefficient of variation termed CV, which can be calculated by dividing the standard deviation by the mean.

$$CV = \sigma_x / \bar{x} \ , \tag{2}$$

where σ_x is the standard deviation and $\bar{x}$ is the mean. The results are shown in the third column of Table 2. The numbers outside the brackets indicate the coefficient of variation of the original data, and the numbers inside the brackets indicate the transformed coefficient of variation.

Given that the amount of student watching is positively proportional to the time the student spends on the courses, we can evaluate the temporal structure of the video viewing sequence to reflect the rationality of the time allocation when students watches the courses. In this regard, Gini coefficient is a suitable indicator, which is shown in columns 6 and 7. It can be learned that the Gini coefficient with respect to viewing time (G-VT) and the Gini coefficient with respect to viewing completion rate (G-VCR) in the same course are relatively close. The two courses with the largest Gini coefficients are EF and MI, and the smallest are LA in G-VT and MM in G-VCR. This result shows that the time structure of the student's consumption ratio of EF and MI is slightly less reasonable than other courses.

Table 2. Measures of temporal structure

	CV	G-VT	G-VCR	ρ
MS	0.744(0.428)	0.399(0.234)	0.392(0.228)	0.999
PE	0.709(0.323)	0.391(0.237)	0.398(0.239)	1.000
LA	0.688(0.400)	0.377(0.223)	0.392(0.234)	1.000
EF	0.770(0.451)	0.410(0.252)	0.405(0.253)	0.990
MM	0.750(0.433)	0.400(0.244)	0.369(0.232)	0.993
MI	0.772(0.441)	0.412(0.246)	0.406(0.234)	0.992
HA	0.780(0.423)	0.397(0.234)	0.388(0.224)	0.948

It is worth noting that CV and Gini characterize the difference in the temporal structure with respect to viewing time and the completion rate of learning, but they show amazing consistency in values. According to literature[25], The Gini coefficient can be approximated as:

$$G = \frac{1}{\sqrt{3}} \frac{\sigma_y}{\bar{y}} \rho(y, r) \ , \tag{3}$$

where σ_y represents the standard deviation, $\bar{y}$ represents the mean, and $\rho(y,r)$ is the correlation coefficient between the student's cumulative viewing and his rank in the population. Both the CV and Gini calculation formulas have the same component, i.e. standard deviation and mean. There is a strong positive correlation between the CV of the transformed data and the G-VT of the original data obtained by the spearman rank correlation test at a significance level of 0.01(coef. = 0.929, p = 0.003); and there is a positive correlation between the CV of the transformed data and the transformed G-VT at a significance level of 0.05(coef. =0.683, p = 0.033). Bringing the mean and standard deviation of Table 2 into the formula, we obtain the ρ values.

5. Course Similarity

5.1 Metric

In order to make the model scalable, we assume that other courses can also be fitted with the gaussian mixture model by choosing the appropriate *k* value.

For each course, we run GMM clustering algorithm and obtain parameters of GMM. The log-likelihood function can be written as:

$$l(\pi, \mu, \sigma) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k \, p\left(x_i; \mu_k, \textstyle\sum_k\right) \right). \tag{4}$$

K-means algorithm is used to initialize model parameters and EM algorithm is used to optimize the parameters.

Given two course samples C^i and C^j, we model them as gaussian mixture model $\Omega^i = \{\varphi_1^i, \varphi_2^i, ..., \varphi_{K_1}^i\}$ and $\Omega^j = \{\varphi_1^j, \varphi_2^j, ..., \varphi_{K_2}^j\}$ where K_1 and K_2 are the number of components of Ω^i and Ω^j respectively. Then, the average similarity of the two course distributions[26] can be computed by:

$$S(\Omega^i, \Omega^j) = 1 - \frac{1}{K_1 K_2} \sum_{h=1}^{K_1} \sum_{l=1}^{K_2} d(\varphi_m^i, \varphi_n^j) \ , \tag{5}$$

$$d(\varphi_m^i, \varphi_m^i) = \frac{1}{8}(\mu_m^i - u_l^i)^T \left(\frac{\Sigma_m^i + \Sigma_n^j}{2}\right)^{-1} (\mu_m^i - u_l^i) + \frac{1}{2} \ln \frac{\left|\frac{\Sigma_m^i + \Sigma_n^j}{2}\right|}{\sqrt{\left|\Sigma_m^i\right| \left|\Sigma_m^i\right|}} \ , \tag{6}$$

where $d(\varphi_m^i, \varphi_m^i)$ is the Bhattacharyya distance that measures the pair-wise similarity of multivariate normal distributions. The bigger of the S value, the similar of the course samples.

5.2 Discussion

The distribution-based course similarity can be applied to personalized course recommendation that addresses the information overload issue by customizing the learning content for students[27]. In previous studies, teachers describe the attributes of courses by analyzing their content, or pre-define corresponding learning goals as the extent to which students would acquire knowledge and skills[28]. The obvious limitation is the lack of a dynamic description of the learning process. Existing course similarity calculation are mainly based on the traditional text mining approaches with a vector space model been constructed according to the knowledge points that each course contains[29]. They mainly suffer from not considering the real-time temporal access patterns towards courses. The courses in the same cluster summarize students' similar learning patterns, which is helpful for assessing the learning process of students.

6. REFERENCES

[1] Z. A. Pardos, "Big data in education and the models that love them," Current opinion in behavioral sciences, vol. 18, pp. 107-113, 2017.

[2] R. S. Baker, Y. Wang, L. Paquette, V. Aleven, O. Popescu, J. Sewall, et al., "Educational data mining: A MOOC experience," Data Mining And Learning Analytics: Applications in Educational Research, pp. 55-66, 2016.

[3] L. C. Liñán and Á. A. J. Pérez, "Educational Data Mining and Learning Analytics: differences, similarities, and time evolution," International Journal of Educational Technology in Higher Education, vol. 12, pp. 98-112, 2015.

[4] N. Buniyamin, U. bin Mat, and P. M. Arshad, "Educational data mining for prediction and classification of engineering students achievement," in 2015 IEEE 7th International Conference on Engineering Education (ICEED), 2015, pp. 49-53.

[5] N. Shahadat and B. Pal, "An empirical analysis of attribute skewness over class imbalance on Probabilistic Neural Network and Naïve Bayes classifier," in 2015 International Conference on Computer and Information Engineering (ICCIE), 2015, pp. 150-153.

[6] H. Guo and H. L. Viktor, "Learning from imbalanced data sets with boosting and data generation: the databoost-im approach," ACM Sigkdd Explorations Newsletter, vol. 6, pp. 30-39, 2004.

[7] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, "Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization," Computers in Human Behavior, vol. 58, pp. 119-129, 2016/05/01/ 2016.

[8] T. Xie, Q. Zheng, and W. Zhang, "Mining temporal characteristics of behaviors from interval events in e-learning," Information Sciences, vol. 447, pp. 169-185, 2018.

[9] I. Molenaar and S. Järvelä, "Sequential and temporal characteristics of self and socially regulated learning," Metacognition and Learning, vol. 9, pp. 75-85, 2014.

[10] H. C. Ratz, "Extreme values from skewed distributions [mathematics education]," IEEE Transactions on Education, vol. 43, pp. 400-402, 2000.

[11] M. Blanca, R. Alarcón, J. Arnau, R. Bono, and R. Bendayan, "Non-normal data: Is ANOVA still a valid option?," Psicothema, vol. 29, pp. 552-557, 2017.

[12] J. W. Osborne, "Improving your data transformations: Applying the Box-Cox transformation," Practical Assessment, Research & Evaluation, vol. 15, pp. 1-9, 2010.

[13] A. Belussi, S. Migliorini, and A. Eldawy, "Detecting skewness of big spatial data in SpatialHadoop," in Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2018, pp. 432-435.

[14] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," Data mining and knowledge Discovery, vol. 13, pp. 335-364, 2006.

[15] Q. Cai, L. Chen, and J. Sun, "Piecewise statistic approximation based similarity measure for time series," Knowledge-Based Systems, vol. 85, pp. 181-195, 2015.

[16] I. Ozken, D. Eroglu, T. Stemler, N. Marwan, G. B. Bagci, and J. Kurths, "Transformation-cost time-series method for analyzing irregularly sampled data," Physical Review E, vol. 91, p. 062911, 2015.

[17] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. Poor, "Mining MOOC Clickstreams: On the Relationship Between Learner Behavior and Performance," arXiv preprint arXiv:1503.06489, 2015.

[18] J. De Boer, P. A. Kommers, and B. De Brock, "Using learning styles and viewing styles in streaming video," Computers & Education, vol. 56, pp. 727-735, 2011.

[19] O. A. Acar and J. van den Ende, "Knowledge Distance, Cognitive-Search Processes, and Creativity: The Making of Winning Solutions in Science Contests," Psychological science, vol. 27, pp. 692-699, 2016/05// 2016.

[20] Y. Chen, B. Zhang, Y. Liu, and W. Zhu, "Measurement and modeling of video watching time in a large-scale internet video-on-demand system," IEEE Transactions on Multimedia, vol. 15, pp. 2087-2098, 2013.

[21] T. Xie, Q. Zheng, W. Zhang, and H. Qu, "Modeling and predicting the active video-viewing time in a large-scale E-learning system," IEEE Access, vol. 5, pp. 11490-11504, 2017.

[22] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," pp. 773-780, 1948/09 1948.

[23] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions," Psychometrika, vol. 52, pp. 345-370, 1987// 1987.

[24] K. P. Burnham and D. R. Anderson, "Multimodel inference - understanding AIC and BIC in model selection," Sociological Methods & Research, vol. 33, pp. 261-304, Nov 2004.

[25] B. Milanovic, "A simple way to calculate the Gini coefficient, and some implications," Economics Letters, vol. 56, pp. 45-49, 1997.

[26] Z. Yu, X. Zhu, H.-S. Wong, J. You, J. Zhang, and G. Han, "Distribution-based cluster structure selection," IEEE transactions on cybernetics, vol. 47, pp. 3554-3567, 2016.

[27] A. Parameswaran, P. Venetis, and H. Garcia-Molina, "Recommendation systems with complex constraints: A course recommendation perspective," ACM Transactions on Information Systems (TOIS), vol. 29, p. 20, 2011.

[28] A. Pawar and V. Mago, "Similarity between learning outcomes from course objectives using semantic analysis, blooms taxonomy and corpus statistics," arXiv preprint arXiv:1804.06333, 2018.

[29] B. Zhao and X. Li, "Course Similarity Calculation Using Efficient Manifold Ranking," in International Conference on Knowledge Science, Engineering and Management, 2015, pp. 421-432.