Towards Evidence-Based Academic Advising Using Learning Analytics

Mariia Gavriushenko(✉), Mirka Saarela, and Tommi Kärkkäinen

Faculty of Information Technology, University of Jyväskylä,
P.O. Box 35, 40014 Jyväskylä, Finland
magavriu@student.jyu.fi, {mirka.saarela,tommi.karkkainen}@jyu.fi

Abstract. Academic advising is a process between the advisee, adviser and the academic institution which provides the degree requirements and courses contained in it. Content-wise planning and management of the student' study path, guidance on studies and academic career support is the main joint activity of advising. The purpose of this article is to propose the use of learning analytics methods, more precisely robust clustering, for creation of groups of actual study profiles of students. This allows academic advisers to provide evidence-based information on the study paths that have actually happened similarly to individual students. Moreover, academic institutions can focus on management and updates of course schedule having an effect of clearly characterized and recognized group of students. Using this approach a model of automated academic advising process, which can determine the study profiles, is presented. The presented model shows the whole automated process, where the learners will be profiled regularly, and where the proper study path will be suggested.

Keywords: Academic advising · Learning analytics
Robust clustering

1 Introduction

To monitor progress of students in their studies is the core measuring activity of an educational institution. The learning goals can be given in a general way, as a list of courses to be passed, or through the credit-based system, which is most commonly used to characterize the study requirements in many learning environments. Credit-based information can be utilized under various student granularities: for an individual student, at the same time enrolled group of students, students with the same major subject, students in one department, or faculty, or university, or country etc. Evidently, aggregated study progress profiles of different groups of students can then be linked to comparisons and assessments, e.g., between different departments, faculties, universities, or national higher education institutions, for the purpose to identify and indicate differently performing educational units.

P. Escudeiro et al. (Eds.): CSEDU 2017, CCIS 865, pp. 1–22, 2018.
https://doi.org/10.1007/978-3-319-94640-5_3

On the level of an individual student, availability of personalized support is very important constituent of a learner's success [41]. When a learner receives punctual personalized advises, the current status of studies and the remaining learning path should be consulted. For this purpose, academic advising (AA) is an iterative collaboration process between student, academic adviser, and academic institution, to tackle especially the student retention. Advisers provide versatile assistance to the students during their studies, making the educational experience relevant and supported.

AA is based on cooperation between adviser, student and institution. It consists of intentional interactions with a curriculum, a pedagogy, and set of students' learning outcomes. AA process is based on relationships between adviser and advisee respectful of the student's concerns. AA activity has been known since 1870 [59]. AA normally starts when a students begin their education in university.

AA normally starts when a student begins his education at the university, and stops when his degree has been completed. Adviser has to make sure that the student is making required studies in time for graduation, that is why academic adviser has to support learner's study path, especially at the beginning of his academic life. Academic adviser should provide for the student his relevant study possibilities and make sure that predefined study plan is being followed. It could depend on organizational culture, especially the stability or dynamicity of the course schedule.

Usually, academic adviser connects with the students through email, social media, web, wikipages etc. for sharing necessary information. However, regular meetings also take place, especially at the beginning of the learning semester or by student's/ adviser's request. For more personalized support, adviser should know the status of the student's studies and when the student is in need of a study advise.

Departments at the universities provide preliminary recommendations to do certain, especially compulsory courses, to proceed normally with the major subject studies. For example, an electronic personal study plan should be prepared with academic adviser at the University of Jyväskylä (JYU). Advising is organized according to the *satellite model* referring to the distributed responsibility of academic units [59]. The *engagement model* between advisee and adviser characterize the principal spirit of counseling [20].

The starting point for a study plan assessment discussion at the JYU is a study plan itself and the completed studies. The actual number of studies that have been made during the academic year is less than recommended. It has been seen especially in computer science. Advising intervention is needed in this situation, because it became common when actual study path deviates from the plans. It might be that student is struggling with something or and he or she needs help. But how does an individual student and especially an individual study adviser know, what is the relation between students' realized study path and that of the peer students? Thus, could we, instead of comparing against the predefined plans, advise students based on evidence from the actual study paths of other similar students?

Therefore, the purpose of this article is to propose an automated AA system based on a learning analytics method [9]. While generally concerned with student models to improve learning, learning analytics methods have shifted from the traditional statistical techniques to the more sophisticated data mining techniques that can also cope with a large amount of unstructured data [21,26,45,50,52]. Here, we propose to use one of these techniques, more precisely robust clustering [6,47], to create groups of actual profiles of students concerning their studies. Such profiles summarize the different typical accumulations of completed studies, increasing the general awareness of the common study flows. That means that individual students can explicitly be linked to their student peers who have a similar study path profile in the same institutional environment. This allows an adviser to plan a possible intervention adaptively for a larger pool of students instead of following each student individually.

Creating general study profiles of students can help departments in their assessment and planning of when (and how) they provide the courses, especially the compulsory ones [47]. By automating general student profiling it is possible to provide essential support for adaptive on-line advising along the lines suggested, for example, in [25,41]. Individual student's perspective, from self-regulated learning and study planning point of view without AA interface, was thoroughly addressed in [4] (see also [5]). By utilizing general student profiles of a higher educational institution we also follow the current trend in learning analytics of not only tackling learning problems of students within a single class- or school-level but to learning analytics problems on an educational organizational level [52]. Such a focus was, for example, recently also addressed by de Freitas et al. [22] and Daniel [16].

The contents of this paper is as follows: after the introduction, we provide a short background on academic advising, we provide a short background on AA in general. Then, in Sect. 3, we describe the robust clustering method and explain how it has been applied in the student data. Also, we present the student profiles and their interpretation. After in Sect. 4 we propose a system model for automated AA with described functions, as well as present a comparison of automated and manual AA.

2 Background

AA is a collaborative process in which adviser and advisee enter a dynamic relationship where adviser helps advisee to enhance the learning experience by helping in making academic decisions [25]. The decision support could be made by analysis of student's records, as well as some external factors like interests, goals, academic capabilities, schedules etc. [42]. Developmental Advising means helping students to define and explore academic and career goals and pathways, as well as to develop problem-solving and decision-making skills; Prescriptive Advising, which is the more traditional advising model, is mainly concentrated on providing the information to the students according to their academic program, progress, academic policies, course selection, etc.; Intrusive Advising refers to contacting with student in critical periods like first year of study before the declaring major, graduation period, or when students are at-risk or they are high-achieving students [42].

Next we briefly summarize a pool of directly AA related work that was identified through a non-systematic search. Our main concern here is to illustrate the strong link between the needs and practices of AA and the general utility of student profiling.

2.1 On AA

Student voices on AA were raised in only some articles. Al-Ansari et al. [2] used questionnaire to study student satisfaction and support-seeking patterns among dental students in Saudi-Arabia. Very low (only 7.6%) primary utility help rate of advisers in the academic matters was encountered. Even if the advisers were available when needed, they were not able to provide the most relevant information, e.g., on important dates and courses. Hence, up-to-date course and timetable information seems to be a prerequisite for AA, which is handled by the in-house developed, integrated study information system *Korppi*[1].

The pedagogical side of AA was also focused rarely. In the work [19] author studied academic advisers through the lens of transformational leadership, i.e. how advisers can create a connection to students that positively influence their study paths (by increasing and inspiring study motivation and engagement/commitment in studies through individual and intellectual consideration). A questionnaire for undergraduate students strengthened the importance of transformational leadership activities in adviser-student communication and collaboration, independently from the student's characteristics. The lack of time for individual counselling efforts that was visible in most of the reviewed articles here was not emphasized in [19]. Dougherty [17] studied academic advisers from those students' perspective who are doing very well in their studies. These students are called high-achieving students. Authors address the need for the investigation of unique characteristics of these students.

Technical support for AA has been considered in many articles. The availability of extensive information on courses to support automatization of AA was emphasized in [7]. The authors proposed course outline data extractor application, which helps in recognizing similar or comparable courses between different institutions, also helping both students and academic advisers to keep track of the variety of topics that (i) have been covered in the completed studies, (ii) should be covered to complete minor or major subject modules or the actual degree. The authors in [41] proposed an integrated knowledge-based framework based on semantic technology that supports computer-based (automatic) e-Advising on the suitable courses for the students. Naturally individual learning history data provide the starting point for the system and, for this purpose, the authors implemented and tested a data integration tool.

[1] https://www.jyu.fi/itp/en/korppi-guide.

The high workload of academic advisers, especially due to individual but many times recurrent handling of basic issues with multiple students in a hurry, was addressed in [25], with the proposition of an intelligent, semantic, web-based application to assist decision making and automatization of repetitive counselling tasks. Core of the system consisted of rule-based inference engine, which mapped student profile with the study program profile and organizational rules, to provide automatic suggestions on the courses to be enrolled in the upcoming semester. In the preliminary evaluation, a positive feedback of the system was obtained, although the main limitation of suitability to only study programs which follow a clear, predefined study path of courses, was recognized. With very similar aims and functionality, another web-based on-line adviser was described in [20]. This system was also evaluated positively when compared to the current advising system. The authors emphasized that such a tool only supports and does not replace a human academic adviser.

Conversational, fully autonomous agent supporting AA dialogs using natural language processing (NLP) were suggested in [35]. The proposed system contained an extensible knowledge base of information and rules on academic programs and policies, course schedules, and a general FAQ. NLP performance of the proposed system was evaluated positively. Also, the similar multi-agent approach was suggested in [61] for AA. This approach helps tackling a dynamic and complex individualized study planning and scheduling problem. As well as in [3] was proposed a decision tree model for AA affairs based on the algorithm C4.5. The output is evaluated based on Kappa measure and ROC area. The main conclusion was made that the difference between the registered and gained credit hours by a student was the main attribute that academic advisers can rely on [2].

As can be concluded, earlier studies have mostly concentrated on research prototypes which focus only on few main components or tool support for existing learning management systems. Taking into account that user modeling is one of the key factors for including personalization into the learning system, many researchers used ontologies for learners' models, because ontologies have many advantages for creation of user models [7,11,25,30,37,41].

Data-mining techniques have also been applied to the learning environments in order to track users' activities, extract their behavior profiles and patterns, and analyze the data for future improvement of the learning results, as well as for identifying types of learners [39]. Mostly, for developing personalized learning plan, researchers used decision tree search, heuristic algorithms, genetic algorithms, item response theory and association rules. Also, many studies used semantic web technologies, neural networks and multi-agent approach. Most of the previous studies on personalized learning path generation schemes have mainly focused on guiding the students to learn in the digital world; i.e. each learning path represents a set of digitalized learning objects that are linked together based on some rules or constraints [38]. While determining such digitalized learning paths, the learning achievements, on-line behaviors or personalized features (such as learning style) of individual students are usually taken into consideration [10,12–14,54].

2.2 On Personalization of Student Support

In general, many researchers have paid attention to developing *e-learning systems with personalization*, and the most common aspect in these system is the *creation of the personalized learning path* for each individual student or group of students. Most of personalized systems consider learner preferences, interests and browsing behaviors, because it will help to provide personalized curriculum sequencing service [27]. In the study [13] authors proposed a personalized e-learning system which is based on Item Response Theory (PEL-IRT). This system is considering course material difficulty and learner ability, to provide individual learning path for learners. Learner's ability estimation was based on an explicit learner's feedback (the answers of learners to the assigned questionnaires). The system appeared mostly like a recommendation system of the courses for the learners. Authors in [27] proposed a genetic-based curriculum sequencing approach and used case-based reasoning to develop a summative assessment. The empirical part indicated that the proposed approach can generate appropriate course materials for learners taking into account their individual requirements. Later, in [11], the authors developed a personalized web-based learning system grounded on curriculum sequencing based on a generated ontology-based concept map, which was constructed by the pre-test result of the learners. Optimization problem for modeling criteria and objectives for automatic determination of personalized context-aware ubiquitous learning path was suggested in [29]. This learning model not only supports learners with alternative ways to solve problems in real-world situations, but also proposes more active interaction with the learners. Authors in [62] proposed Decision Support System for student advising based on decision tree for an automated program planning and scheduling. The proposed approach takes into account prerequisite rules, the minimum time (minimum number of terms), and the academic recommendations. The adaptive course sequencing for personalization of learning objectives was suggested in [30] using neural networks, self organizing maps and the back-propagation algorithm.

A very closely related work to ours was reported in [53]. Authors proposed a case-based reasoning paradigm which is based on the assumption that similar students will have similar course histories. The system used the experience and history of graduated students in order to propose potential courses for the students. Unfortunately, this approach required matching between students' histories. Also, similar case-based reasoning was used by [40] for developing a recommendation system for a suitable major to students based on comparison of the student information and similar historical cases.

As reviewed, many suggestions for intelligent software and information system support of AA have been given. Many studies describe the creation of intelligent learning systems that can make a curriculum sequencing more flexible for providing students with personalized and adaptive study support services

[8,23,36,36,43,57]. Universities are more and more looking into developing self-service systems with intelligent agents as an addition or replacement for the labor-intensive services like AA. For example, The Open University of Hong Kong has developed an intelligent on-line system that instantly responds to enquiries about career development, learning modes, program/course choices, study plans, and graduation checks [37].

However, the institutional starting point concerning available digital information, especially for the web-based systems that have been proposed, seems to vary a lot. Some systems start and focus on providing easy access to course and degree requirements information whose availability is to be assured first. On the other hand, we might start from the situation where we can readily access most of the relevant data: (i) course information with basic contents, learning goals, assessment methods, acceptance criteria, schedule and location, teachers and lecturers etc.; (ii) individual, anonymous study records on passed courses and completed studies. (Note that reliable information on student admission is currently not directly available in the organization under consideration).

3 Creation of Study Path Profiles Using Robust Clustering

3.1 Data

To illustrate the proposed approach, we utilized the real study records of the Bachelor (B.Sc.) and Master (M.Sc.) students majoring in Mathematical Information Technology at the University of Jyväskylä (JYU/MIT). The MIT programs at JYU are comparable to a major in Computer Science at other universities. The information technology administration at JYU has recently created a data warehouse of passed courses from all the students. On one hand, these data can be utilized by the departments. On the other hand, the electronic study plan system does not provide a direct interface for larger student groups. That means that both from accessibility and evidence-basedness points of view, we focused on analyzing the real study log of the passed courses covering the four calendar years 2012–2015. The log was anonymized by keeping the student IDs as keys. It should be noted that students can start their studies either in the beginning of September (autumn term) or January (spring term). Hence, the original study registry log included a heterogeneous set of B.Sc. and M.Sc. students who had started their studies either before 2012 or in the beginning of spring or autumn terms during 2012–2015.

The whole study log contained 15370 passed courses by 1163 different students on 1176 different course IDs. Altogether, there were 942 male students (81%) in the data, who on average completed 59.9 ECTS. Only 221 female students (19%) of the students were female and they had completed on average 57.0 ECTS. Since at least 180 ECTS are needed for a B.Sc. and an additional 120 ECTS for a M.Sc. degree, we conclude that most of the students in the log were either in the beginning of their studies or—as already reported for the previous student data in [47]—progressing very passively and slowly. Both, the gender distribution as well as the slow progression, assemble rather typical situations in Computer Science programs [15, 34, 44, 58].

Figure 1 shows the discrete density distribution of the size of the passed courses. According to the figure, 5 ECTS and 3 ECTS are the two most common sizes of the courses, the former covering around 30% of the studies. Moreover, there are a lot of small courses (1–6 ECTS) with the exception of the M.Sc. thesis, which is worth 30 ECTS. Teaching in JYU is organized into four periods during one academic year (plus the summer semester) in such a way that a course of circa 5 ECTS can fit to one period. Since the passed courses represent both major and minor subject studies, we argue here that the division of the overall learning objects as courses is not optimal. During the course of writing this article, we also discovered that the instructions of JYU for preparing the next curriculum for 2017–2019 include a strong recommendation to decrease the number of courses with only a few credits. Thus, this observation already provides an example on how summarization of study log data produces visibility and feedback to the organization.

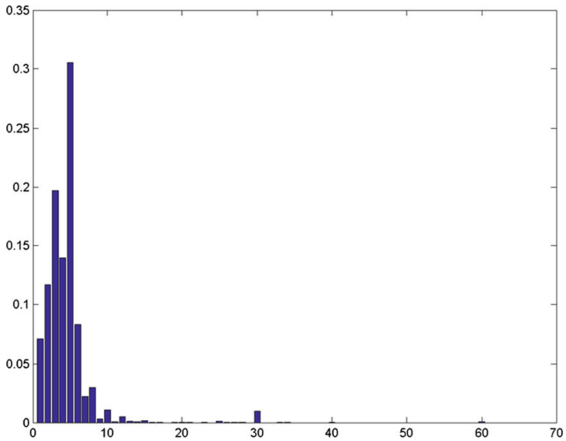

Fig. 1. Probability of size of a course [24].

In a next step, we aggregated how many credits per semester each student completed. This was performed similarly as in [47]: each calendar year was divided into two semesters, the spring term (from January to June) and the autumn term (from July to December). As explained above, there are three academic semesters, spring, summer (June–August), and autumn, where spring and autumn are further divided into two periods. However, since usually only a few courses are completed during the summer (this is illustrated in [47]), it was reasonable to divide the calendar year into only two parts for further analysis.

The semesterwise study accumulation represents the overall study activity: how many credits were made by a student per semester. Our second main concern, directly related to AA, is to study how effectively (i.e., when) the compulsory part of the B.Sc. major studies has been made. In our case, during 2012–2015, there were ten different compulsory courses of interest, for which both autumn and spring term starters had their separate study recommendations. Some, but not all, compulsory courses were provided more than once in an academic year. Hence, as the second transformation of the study log, we created a student-wise indication of that semester when one of the ten compulsory bachelor courses was passed. Naturally, the counting of semesters started from that one when the student was enrolled to the studies. This yields to 10 integer variables of range $[1, 8]$.

In what follows, we profile, analyze and compare two students cohorts: those who started their studies in the beginning of the autumn term 2012 (A2012) or 2013 (A2013). Hence, for A2012 we end up with 8 and for A2013 with 6 integer variables representing the aggregated amount of credits on half-a-year scale. Since the students have progressed in their studies very differently and many of them have not been active during all the semesters of interest, both of the data sets are very sparse containing a lot of missing values [47]. This is the key property that is taken into account in the profiling approach that is described next.

3.2 Robust Clustering Method

As already explained, our goal is to assist the academic advisers by recommending suitable courses for students based on passed courses of (possibly more advanced) students with similar study path. For this, we need to identify general profiles of similar students and this is, precisely, the purpose of clustering. Partitional (or representative-based [63]) clustering seems to be the right family of clustering methods to choose from because it assigns each observation to exactly one cluster, which is represented by its most characteristic point, the cluster centroid, which represents the common profile. Within a cluster, distances of observations to the prototype determine the most typical or representative members of a cluster. Thus, instead of following many different student profiles, the academic adviser can just follow the most common profiles to get an overview of the whole cohort.

Generally, partitional based clustering algorithms consist of an initialization step, in which the initial centroids of each cluster are generated, and iterations of two steps where (i) each observation is assigned to its closest centroid, and (ii) the centroid of each cluster is recomputed by utilizing all observations assigned to it. The algorithm stops when the centroids remain the same over two iterations. The most popular and most applied partitional clustering algorithm is the *k-means* [31], also in learning analytics studies [52]. This algorithm works very well for full and approximately normally distributed data since the sample mean is the most efficient estimator for samples that are drawn from the normal distribution. However, the sample mean is highly sensible to all kinds of outliers [28] as well as missing values, which can be characterized as special types of outliers. Since our data is—as explained in Sect. 3.1–very sparse, this clustering approach would not be suitable here. Furthermore, for a non-symmetric (skewed) distribution, the sample mean is not necessarily the most efficient estimator and other location estimates might be preferable [50,55]. Moreover, as more thoroughly explained in [47], the quantization error for the integer-type variables like here has uniform not Gaussian distribution.

The spatial/geometric median is a robust nonparametric location estimate, which remains reliable even if half of the data is contaminated [33,55]. Mathematically, the spatial median is the Weber point that minimizes the (nonsquared) sum of the Euclidean distances to a group of given points $\{\mathbf{x}_i\}, i = 1, 2, \ldots n$:

$$\arg\min_{\mathbf{c}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{c}\|. \tag{1}$$

Although the basic concept is easily understood and has been extensively discussed in the literature (albeit under various names, see [18]), its computation is known to be difficult.

In [6], the difficulty of computing the spatial median during partitional clustering was solved with the SOR (Sequential Overrelaxation) algorithm (see [6] for details). Moreover, in the implementation of the resulting *k-spatial-medians* clustering algorithm, only the available (i.e., the not-missing) data is taken into account when the centroid is recomputed.

To sum up, all of these above discussed properties—most importantly, the robustness to missing data and the fact that every cluster is represented by a centroid—make the *k-spatial-medians* clustering very suitable for creating student's general study profiles. The fact that such a clustering approach works very well for sparse educational data has been previously shown in [46,48,49,51]. The initialization of the robust clustering method was realized similarly as in [49]: We started with multiple repetitions of *k-means* for the complete data – without missing values – and then, applied *k-spatial-medians* to the best of those results.

Concerning the second representation of the study log, we note that the information whether a mandatory course was passed by a particular student in a certain semester is a binary one. This means that to find the semester-wise co-occurrences one could apply association rule mining, especially the famous Apriori-algorithm [1]. Hence, we note that use of *k-spatial-medians* clustering with the proposed encoding provides a scalable and—since we do not need to binarize variables—also an easier and faster alternative to association rule mining with very similar support for the educational knowledge discovery.

3.3 Clustered Student Profiles

Similarly to the earlier work in [46–48,60], we apply four different internal cluster validation indices to determine the number of clusters: Knee Point (KP) of the clustering error, Ray-Turi (RT), Davies-Bouldin (DB), and Davies-Bouldin∗ (DB∗). All the computations here were carried out in the Matlab-environment, using own implementations of all the algorithms.

From the two student groups A2012 and A2013, we include in clustering only still active students, i.e. those who have made credits during the autumn term 2015 (the last one analyzed). Furthermore, we restrict ourselves to those students for whom over half of the variables are available [55]. This means that the 47 analyzed students in A2012 have made studies during at least four out of the seven possible semesters (including the last one) and the 76 students in A2013 at least in three out of the five semesters. Because of the anonymity, we obtained further assistance in relation to the metadata and interpretation of the clusters from the Study Amanuensis of the Department [56].

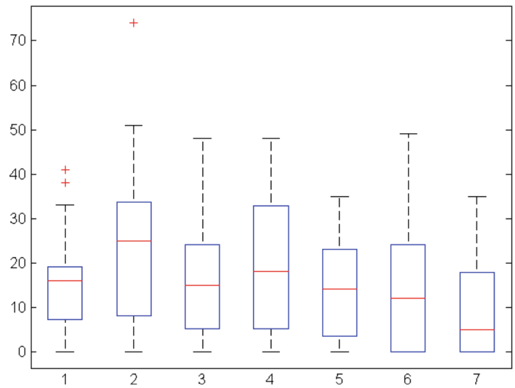

Fig. 2. Boxplot for A2012 [24].

A2012. The boxplot in Fig. 2 shows the large variability in the study accumulations both within semesters and between semesters. We see the larger accumulations in the spring terms during the first two years, and a slightly decreasing overall trend after that. There are always exceptional students who have made much more studies than their peers.

KP, DP, and DP∗ indicated four clusters and RT had also local minimum there, so we choose to analyze four different general study progress profiles. The profiles for A2012 are depicted in Fig. 3, where the size of the cluster is given in the top-right corner. The profiles are sorted in the ascending order with respect to the total number of credits.

The main group of 21 students in the first cluster illustrate a potential start of the studies in the first year, with strong passivation after that. They have obtained prototypically 65 ECTS until the end of 2015. Based on [56], by a closer look on the 8 students from the cluster closest to the centroid, these are all older B.Sc. and M.Sc. male students (born before 1990). They are either distant students studying while working or have completely chosen to change their orientation from an earlier occupation and already finished degree. The difficulties in studies and reasons of such a behavior, for a similar adult student profile, were thoroughly discussed in the earlier work [32] from the same context (department) than here.

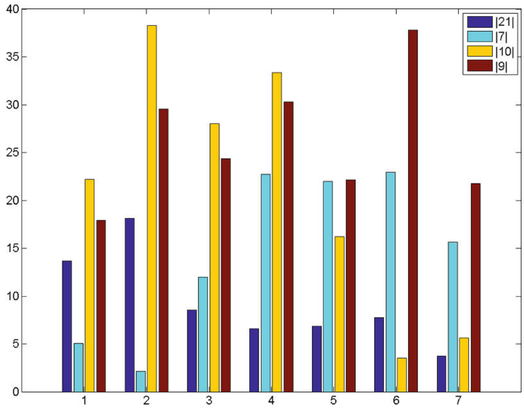

Fig. 3. Credit accumulation prototypes for A2012 [24].

The second group of only 7 students, who generally obtained 103 ECTS, shows opposite behavior: very slow start in the first year, activating to an appropriate level then. Three most characteristic students here were young males, who were involved in the military service during the first study year. This complete explains the observed behavior.

The third group of 10 students, who generally obtained 147 ECTS, did their studies very actively for the first 4–5 semesters. Analysis of the three most characteristics students revealed two young and one older male students who either took job or became active in student organizations during the third year of the studies.

The fourth profile with 9 students, altogether 184 ECTS in general, illustrates that a good start on the study activity carries over the semesters. Three mostly characteristics students were again all males, one M.Sc. student and two B.Sc. students. Note that similar finding on the importance of active start in an individual course level was given in [47].

Students who are mostly in need of AA are the ones in the first cluster. They can be identified either in the beginning of their studies or after the second semester, because even if still making studies, their accumulation is much less than in the third and fourth cluster. Their characterization also suggests the department to rethink the study entrance criteria.

A2013. For A2013 all cluster indices suggested three profiles, which are illustrated in Fig. 5. This and the fact that there are now one profile less than in A2012 suggests more stable organization of the curriculum. Also the boxplot in Fig. 4 supports such finding, especially showing smaller variability in the obtained credits between the autumn and spring terms compared to A2012 in Fig. 2.

Student group in the need of intrusive AA consists of those 23 students with smallest accumulation of credits. These students start and continue very slowly in their studies, although the level of activity was increasing in the fourth and fifth semesters. Their general ECTS accumulation after five semesters was 44 ECTS. Analysis of the five most representative students revealed two older male students (birth year before 1990), two males with indications of military service, and a female student. According to [56], especially the younger students showed signs of low self-regulation during AA sessions.

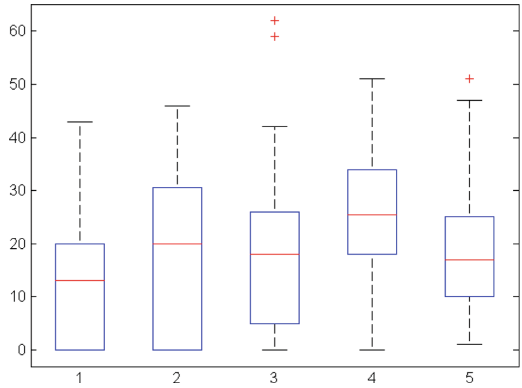

Fig. 4. Boxplot for A2013 [24].

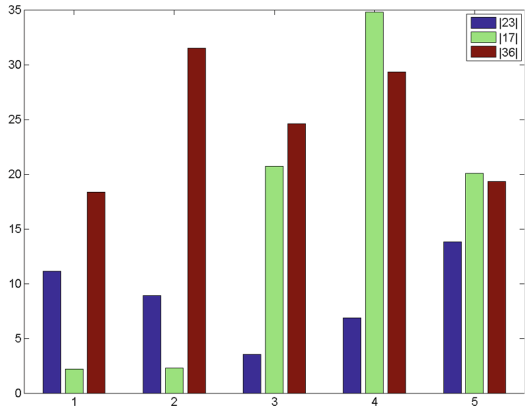

Fig. 5. Credit accumulation prototypes for A2013 [24].

The second profile of 17 students, completing typically 80 ECTS, showed similar behavior to the second profile in A2012: the minimal first year is raised to a good level of study activity later. A closer look on the five most representative students showed young, two female and three male students. Four of these had identified themselves as a non-active student during the first study year, again mostly due to the military service of the young male students.

The third profile of 36 most active students, accomplishing 123 ECTS typically, showed similar overall behavior than the fourth profile in A2012. The first semester is slightly smaller but then the study path proceeds in the desired way. Recapitulation of the meta data of five most representative students showed five male students, of whom three were oriented towards game programming and development - the most recent study line of the department.

We note that even if the boxplot in Fig. 4 indicated more stable study path with respect to autumn and spring semesters, the two profiles of truly active students still illustrated larger study accumulations in the spring than in the autumn. These findings are, however, mostly explained by the longer calender time for the two periods in the spring term compared to the autumn term – a general peculiarity of the Finnish higher education system.

The similarities and differences between the two sets of profiles just discussed emphasize the importance of the use of evidence-based information in AA. On one hand, there are repetitive profiles of students proceeding in their studies well or slowly. The latter ones needs to be detected and supported in an intrusive manner in AA. The home department responsible for major subject studies and the other departments providing minor subject studies should be informed about the found hindrances of the study paths. In the case analyzed here, there is a clear change of study accumulation profiles from A2012 to A2013, which suggests that the organization of courses, the capabilities of students, and/or their support through AA have improved in the educational organization under study.

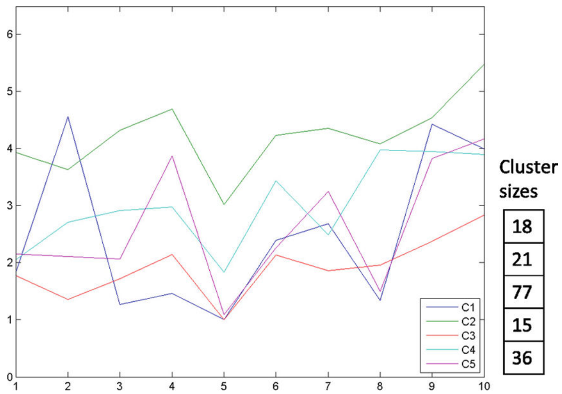

Fig. 6. Mandatory B.Sc. course prototypes for active MIT students 2012–2014.

The third clustering result related to pass semester of the mandatory courses is given in Fig. 6. By far the largest cluster C3 with 77 students presents a favorable pattern where all mandatory courses are made during the early study semesters. This suggests that the organizational arrangements to offer equal study possibilities for the students starting either in autumn or spring semester are working as planned for most of the students. However, cluster C4 with 15 students shows a kind of one-semester-delayed pattern of passing the courses. Similarly to the previous two profiles, also this clustering result encapsulates the cluster C2 with 21 students, who start slow/late but then proceed desirably. For a group of 18 students in C1, the second mandatory course is postponed very late, and for a group of 36 students in cluster C5, the fourth mandatory course is passed later than usual. Altogether, patterns in clusters C2, C4, and C5 indicate need of further efforts in both course offerings or arrangements and AA.

4 Proposition of a System Model

As shown, it is important to follow the actual progress of the students in their studies. There might be no need for an advising intervention, but if so, one should automatically notify the students and the study counselling on the deviations in the study path. The problems of not passing courses and not following suggested study plans usually also call for organizational considerations whether learner ability and the difficulty level of the recommended curriculum are matched to each other properly [27].

4.1 Proposed System Architecture

This subsection describes the novel system architecture for AA and automatic feedback based on recognized student group profiles which are obtained by using clustering. We also present an overview of the AA process as both manual and automated process.

The architecture of the proposed system's model for the AA is presented in Fig. 7. The system has two main databases: learner profile database and curriculum database. Learner profile database stores learner's data about studies, assessment results, timetables of completed studies, etc. Curriculum database stores information about compulsory courses, other courses, timetables, etc. The AA system's part consists of several blocks like linking individual students to their peers with similar study path profile together with the recommendation block and planning block.

Based on the system architecture, the details of system's main functionality read as follows:

1. Collection of learner's personal information.
2. Collection of information about the courses and completed studies.
3. Creation of study progress profiles along the lines of Sect. 3.
4. Linking the individual student to student peers with similar study profile in the institutional environment.
5. Student's progress check. If student is linked to a profile requiring intrusive advising, inform the adviser and the student by providing the interpreted study profile to support the communication and problem solving.

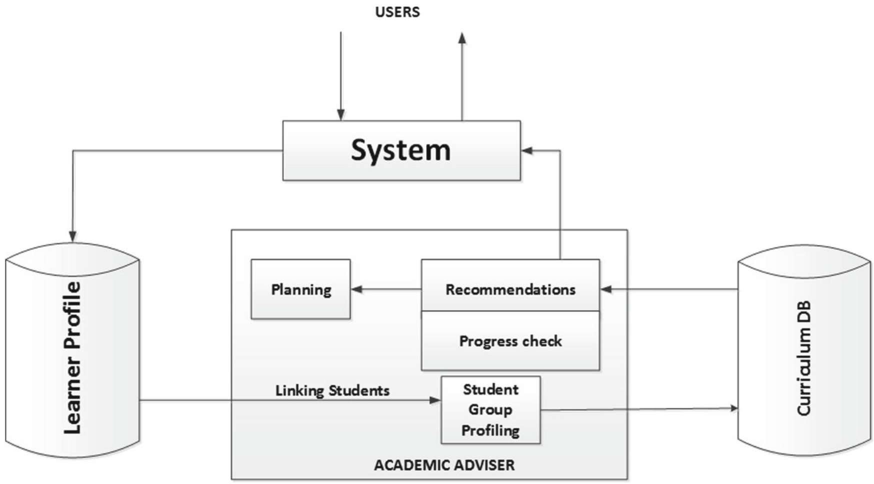

Fig. 7. Architecture of the academic adviser [24].

6. Modification of the study plan on recommended courses and their timetables by taking into account the evidence related to the identified study profile.
7. Planning and realization of an intrusive profile intervention adaptively for a larger pool of students.

Data collection related to the system is, naturally, all the time active. The evidence-based study profiles can and should be recomputed on regular basis. A natural suggestion would be to do this after the studies made during the previous semester have been stored and become available for clustering.

4.2 Automated Academic Advising Process

The automated process of AA, related to the system's architecture and main functionality as described above, is presented in Fig. 8. The given proposition allows manual control of the continuous advising activity for every learner individually, or the more automated process where the role of the advisers is shifted to the higher level of abstraction. The difference on the level of learner's life cycle between these two use cases is depicted in Fig. 9. The automated process is highlighted with the red color in the figure. In the automated scenario, the responsible persons of the study organization only provide policies, planning and regulations. This can reduce the responsibility for the daily routine work and could help to provide recommendations for a larger pool of students rather than for the each individual learner.

The work-flow related to Fig. 8 reads as follows: The learner is choosing the study program of an educational institution. After that he or she chooses with AA the proper courses which are related to the chosen program and creates a study plan. Information about the student, the required courses and progress

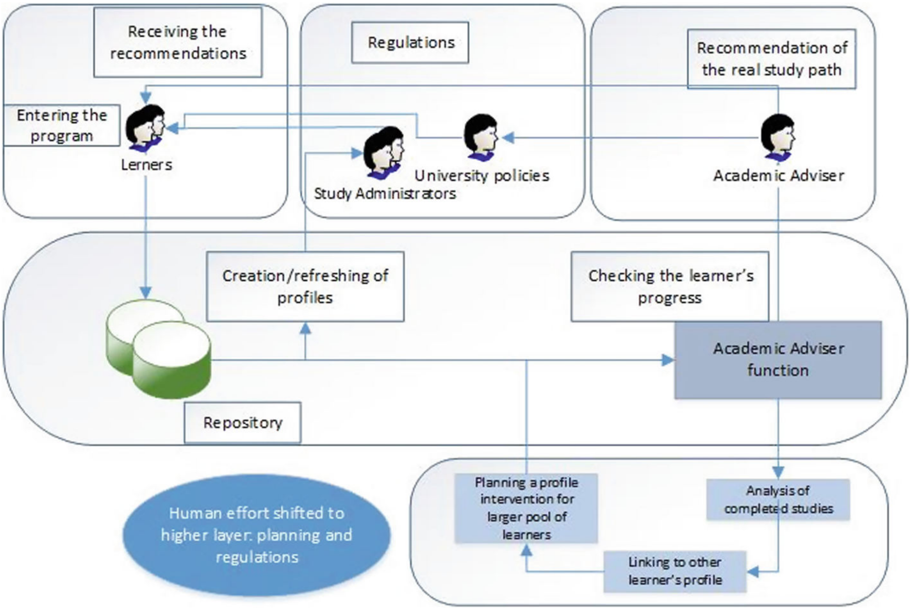

Fig. 8. Automated academic advising process [24].

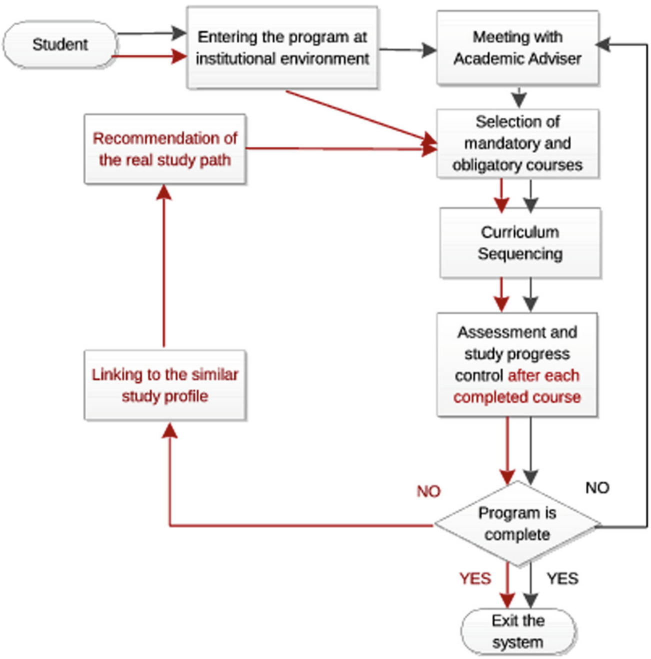

Fig. 9. Comparison of the manual and automated processes of learner's study life cycle [24]. (Color figure online)

in them is stored in the database and is automatically changed/refreshed after each passed course. After passing several courses, system can attach a student to a group of students with similar, actual study path. If learners are doing well, evidence-based determination and communication of this during advising encourages them to continue like that. If they are attached to a profile which does not progress with the studies as expected, the system can identify this early and provide intrusive AA support for both the advisor and the students in question.

The proposed automated mechanism solve an important problem of improving and providing academic advising, because more and more students should receive guidance with their study plans before graduating. This system will help to plan when and how to provide the courses, especially the compulsory ones, as well as to plan a profile intervention adaptively for a larger pool of students, which will reduce the human effort of academic advising.

The proposed architecture and model of the system are intended for a development phase to prototype the whole automated process, where the learners will be profiled regularly, and where the proper study path will be presented, as well as deviating learners detected. The model of the AA system will have automated process of study path recommendation. This system will help to plan when and how to provide the courses, especially the compulsory ones, as well as to plan a profile intervention adaptively for a larger pool of students, which will reduce the human effort of AA.

5 Conclusions

The topic of AA has been the subject of many research efforts, but in academic practice, more often than not, AA is still not reaching the desired level of effectiveness and efficiency. There is a need for a proper personalized AA, and it is very important in academic life. It is decisive that learner should receive proper advising – poor or no advising is known to have a negative effect on the progress in studies [2].

In this work a compact literature review about AA, mostly focusing on Automated Academic Advising and Intelligent Academic Advising was presented. This paper aims at improving the practice of AA by proposing a model of an automated AA system. The key component of this model is based on learning analytics, more precisely robust clustering, which is especially suitable for data with missing values. This method helps in detecting student groups (clusters) based on the shared study paths. Study path is considered to be a sequence of courses a student has taken and completed during his/her studies.

A system architecture towards the development of an automated system for the academic advising process was presented. The system is in the development stage. However, the discussion and comparison of the steps of the manual and the proposed automated academic advising process was shown. The system will recommend a proper study path according to profiled learners, as well as deviating learners will be detected and will be supported by academic advisers. This automated process will help to plan interventions for a larger pool of students and that will reduce human effort of AA. Moreover, academic institutions can focus on management and updates on course schedule knowing the a clearly characterized groups of students. Note that even if the sample groups of students that were profiled here were very small, the used method is scalable to hundreds of thousands of students [48].

The identified student groups and the corresponding study trajectories are based now only on the students' progress – the number of credits they are earning, but not on any other information that would characterize the courses that form the trajectory. e.g. subject matter of the course, student's achievement, if it is obligatory or elective etc. That is why by continuing the development of this work, we could take into account analysis which is not only based on the number of credits, but also on student's majors, minors, subject matter interests to consider the study paths with higher granularity than per semester. Moreover, future work should include in the main functionality of the proposed system an automated notification for the academic advisers about students and their progress, with the interpretation of needs to modify and re-plan the study path. Furthermore, the proposed system should be properly evaluated. A number of users will evaluate it: whereas interviews can help to qualitatively estimate the results or questionnaires to find out if the system architecture was designed suitably, according to users needs, and if the recommended learning path was appropriate for the student.

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: 1995 Proceedings of the Eleventh International Conference on Data Engineering, pp. 3–14. IEEE (1995)
2. Al-Ansari, A., El Tantawi, M., AbdelSalam, M., Al-Harbi, F.: Academic advising and student support: help-seeking behaviors among Saudi dental undergraduate students. Saudi Dental J. **27**(2), 57–62 (2015)
3. Al-Sarem, M.: Building a decision tree model for academic advising affairs based on the algorithm C 4–5. arXiv preprint arXiv:1511.04026 (2015)
4. Auvinen, T.: Educational technologies for supporting self-regulated learning in online learning environments. Aalto University (2015)
5. Auvinen, T., Paavola, J., Hartikainen, J.: STOPS: a graph-based study planning and curriculum development tool. In: Proceedings of the 14th Koli Calling International Conference on Computing Education Research, pp. 25–34. ACM (2014)
6. Äyrämö, S.: Knowledge Mining Using Robust Clustering. Jyväskylä Studies in Computing, vol. 63. University of Jyväskylä, Jyväskylä (2006)
7. Biletskiy, Y., Brown, J.A., Ranganathan, G.: Information extraction from syllabi for academic e-Advising. Expert Syst. Appl. **36**(3), 4508–4516 (2009)
8. Brusilovsky, P.: Adaptive educational systems on the world-wide-web: a review of available technologies. In: Proceedings of Workshop "WWW-Based Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS 1998), San Antonio, TX (1998)
9. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. Int. J. Technol. Enhanc. Learn. **4**(5–6), 318–331 (2012)

10. Chen, C.M.: Intelligent web-based learning system with personalized learning path guidance. Comput. Educ. **51**(2), 787–814 (2008)
11. Chen, C.M.: Ontology-based concept map for planning a personalised learning path. Br. J. Educ. Technol. **40**(6), 1028–1058 (2009)
12. Chen, C.M.: Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. Expert Syst. Appl. **36**(5), 8816–8829 (2009)
13. Chen, C.M., Lee, H.M., Chen, Y.H.: Personalized e-learning system using item response theory. Comput. Educ. **44**(3), 237–255 (2005)
14. Chen, C.M., et al.: Ontology-based concept map for planning personalized learning path. In: IEEE International Conferences on Cybernetics & Intelligent Systems (2008)
15. Cohoon, J.M., Aspray, W.: Women and Information Technology: Research on Underrepresentation, vol. 1. The MIT Press, Cambridge (2006)
16. Daniel, B.: Big data and analytics in higher education: opportunities and challenges. Br. J. Educ. Technol. **46**(5), 904–920 (2015)
17. Dougherty, S.B.: Academic advising for high-achieving college students. High. Educ. Rev. **4**, 63–82 (2007)
18. Drezner, Z., Hamacher, H.W.: Facility Location: Applications and Theory. Springer, Heidelberg (2001). https://doi.org/10.1007/978-3-642-56082-8
19. Drozd, D.S.: Student preferences for academic advisors as transformational leaders. Ph.D. thesis, Texas A&M University (2010)
20. Feghali, T., Zbib, I., Hallal, S.: A web-based decision support tool for academic advising. Educ. Technol. Soc. **14**(1), 82–94 (2011)
21. Ferguson, R.: Learning analytics: drivers, developments and challenges. Int. J. Technol. Enhanc. Learn. **4**(5–6), 304–317 (2012)
22. de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., Arnab, S.: Foundations of dynamic learning analytics: using university student data to increase retention. Br. J. Educ. Technol. **46**(6), 1175–1188 (2015)
23. Fung, A., Yeung, J.: An object model for a web-based adaptive educational system. In: Proceedings of the IFIP International Conference on Educational Use of Technologies (ICEUT 2000) (2000)
24. Gavriushenko, M., Saarela, M., Kärkkäinen, T.: Supporting institutional awareness and academic advising using clustered study profiles. In: Proceedings of the 9th International Conference on Computer Supported Education: CSEDU, vol. 1, pp. 35–46. INSTICC, ScitePress (2017)
25. Henderson, L.K., Goodridge, W.: AdviseMe: an intelligent web-based application for academic advising. Int. J. Adv. Comput. Sci. Appl. **6**(8), 233–243 (2015)
26. Hershkovitz, A., Knight, S., Dawson, S., Jovanović, J., Gašević, D.: About "learning" and "analytics". J. Learn. Anal. **3**(2), 1–5 (2016)
27. Huang, M.J., Huang, H.S., Chen, M.Y.: Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. Expert Syst. Appl. **33**(3), 551–564 (2007)
28. Huber, P.J.: Robust Statistics. Springer, Heidelberg (2011)
29. Hwang, G.J., Kuo, F.R., Yin, P.Y., Chuang, K.H.: A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning. Comput. Educ. **54**(2), 404–415 (2010)
30. Idris, N., Yusof, N., Saad, P., et al.: Adaptive course sequencing for personalization of learning path using neural network. Int. J. Adv. Soft Comput. Appl. **1**(1), 49–61 (2009)

31. Jain, A.K.: Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. **31**(8), 651–666 (2010)
32. Kaihlavirta, A., Isomöttönen, V., Kärkkäinen, T.: A self-ethnographic investigation of continuing education program in engineering arising from economic structural change. Stud. Contin. Educ. **37**(1), 99–114 (2015)
33. Kärkkäinen, T., Saarela, M.: Robust principal component analysis of data with missing values. In: Perner, P. (ed.) MLDM 2015. LNCS (LNAI), vol. 9166, pp. 140–154. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21024-7_10
34. Kinnunen, P., Marttila-Kontio, M., Pesonen, E.: Getting to know computer science freshmen. In: Proceedings of the 13th Koli Calling International Conference on Computing Education Research, Koli Calling 2013, pp. 59–66. ACM, New York (2013). https://doi.org/10.1145/2526968.2526975
35. Latorre-Navarro, E.M., Harris, J.G.: An intelligent natural language conversational system for academic advising. Editor. Preface **6**(1) (2015)
36. Lee, M.G.: Profiling students' adaptation styles in web-based learning. Comput. Educ. **36**(2), 121–132 (2001)
37. Leung, C.M., Tsang, E.Y., Lam, S., Pang, D.C.: Intelligent counseling system: a 24 x 7 academic advisor. Educause Q. **33**(4), n4 (2010)
38. Liu, C.L., Wu, S., Chang, M., Heh, J.: Guiding students to do remedial learning in school campus with learning objects' spatial relations. In: 16th International Conference on Computers in Education Conference, pp. 249–256. Citeseer (2008)
39. Minaei-Bidgoli, B.: Data mining for a web-based educational system. Ph.D. thesis, Michigan State University (2004)
40. Mostafa, L., Oately, G., Khalifa, N., Rabie, W.: A case based reasoning system for academic advising in Egyptian educational institutions. In: 2nd International Conference on Research in Science, Engineering and Technology (ICRSET 2014), pp. 21–22, March 2014
41. Nguyen, T.B., Nguyen, D.N., Nguyen, H.S., Tran, H., Hoang, T.A.D.: An integrated approach for an academic advising system in adaptive credit-based learning environment. J. Sci. Nat. Sci. Technol. **24**, 110–121 (2008)
42. Noaman, A.Y., Ahmed, F.F.: A new framework for e academic advising. Procedia Comput. Sci. **65**, 358–367 (2015)
43. Papanikolaou, K.A., Grigoriadou, M., Magoulas, G.D., Kornilakis, H.: Towards new forms of knowledge communication: the adaptive dimension of a web-based learning environment. Comput. Educ. **39**(4), 333–360 (2002)
44. Papastergiou, M.: Are computer science and information technology still masculine fields? High school students' perceptions and career choices. Comput. Educ. **51**(2), 594–608 (2008)
45. Rogers, T.: Critical realism and learning analytics research: epistemological implications of an ontological foundation. In: Proceedings of the 5th International Conference on Learning Analytics and Knowledge, pp. 223–230. ACM, New York (2015). https://doi.org/10.1145/2723576.2723631
46. Saarela, M., Kärkkäinen, T.: Discovering gender-specific knowledge from Finnish basic education using PISA scale indices. In: Proceedings of the 7th International Conference on Educational Data Mining, pp. 60–68 (2014)
47. Saarela, M., Kärkkäinen, T.: Analysing student performance using sparse data of core bachelor courses. JEDM-J. Educ. Data Min. **7**(1), 3–32 (2015). http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM056
48. Saarela, M., Kärkkäinen, T.: Do country stereotypes exist in PISA? A clustering approach for large, sparse, and weighted data. In: Proceedings of the 8th International Conference on Educational Data Mining, pp. 156–163 (2015)

49. Saarela, M., Kärkkäinen, T.: Weighted clustering of sparse educational data. In: Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015), pp. 337–342 (2015)
50. Saarela, M.: Automatic Knowledge Discovery from Sparse and Large-scale Educational Data: Case Finland. Jyväskylä Studies in Computing, vol. 262. University of Jyväskylä, Jyväskylä(2017)
51. Saarela, M., Hämäläinen, J., Kärkkäinen, T.: Feature ranking of large, robust, and weighted clustering result. In: Kim, J., et al. (eds.) PAKDD 2017. LNCS, vol. 10234, pp. 96–109. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57454-7_8
52. Saarela, M., Kärkkäinen, T.: Knowledge discovery from the programme for international student assessment. In: Peña-Ayala, A. (ed.) Learning Analytics: Fundaments, Applications, and Trends. SSDC, vol. 94, pp. 229–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52977-6_8
53. Sandvig, J., Burke, R.: AACORN: a CBR recommender for academic advising. Technical report TR05-015, DePaul University (2005)
54. Schiaffino, S., Garcia, P., Amandi, A.: eTeacher: providing personalized assistance to e-learning students. Comput. Educ. **51**(4), 1744–1754 (2008)
55. Sprent, P., Smeeton, N.C.: Applied Nonparametric Statistical Methods. CRC Press, Boca Raton (2016)
56. Study Amanuensis: Private Communication, 2 November 2016
57. Tang, T.Y., McCalla, G.: Smart recommendation for an evolving e-learning system: architecture and experiment. Int. J. eLearning **4**(1), 105 (2005)
58. The Daily Telegraph: University degree subjects with the highest dropout rates (2017). http://www.telegraph.co.uk/education/educationpicturegalleries/11002595/University-degree-subjects-with-the-highest-dropout-rates.html?frame=2824912. Accessed 01 Mar 2017
59. Tuttle, K.N.: Academic advising. New Dir. High. Educ. **2000**(111), 15–24 (2000)
60. Wallden, L.J.: Kansainvälisten koulutusarvioiden vertailu koulutuksellisen tiedonlouhinnan keinoin. University of Jyväskylä, Jyväskylä Studies in Computing (2016)
61. Wen, F.L.S.L.D., McGreal, F.Z.K.R.: e-Advisor: a multi-agent system for academic advising. Int. J. Adv. Comput. Sci. Appl. **6**(8) (2015)
62. Werghi, N., Kamoun, F.K.: A decision-tree-based system for student academic advising and planning in information systems programmes. Int. J. Bus. Inf. Syst. **5**(1), 1–18 (2009)
63. Zaki, M.J., Meira Jr., W.: Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, Cambridge (2014)