

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Gavriushenko, Mariia; Saarela, Mirka; Kärkkäinen, Tommi

Title: Supporting Institutional Awareness and Academic Advising using Clustered Study Profiles

Year: 2017

Version:

Please cite the original version:

Gavriushenko, M., Saarela, M., & Kärkkäinen, T. (2017). Supporting Institutional Awareness and Academic Advising using Clustered Study Profiles. In P. Escudeiro, G. Costagliola, S. Zvacek, J. Uhomobhi, & B. M. McLaren (Eds.), CSEDU 2017 : Proceedings of the 9th International Conference on Computer Supported Education. Vol. 1 (pp. 35-46). SCITEPRESS Science And Technology Publications.
<https://doi.org/10.5220/0006252300350046>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Supporting Institutional Awareness and Academic Advising Using Clustered Study Profiles

Mariia Gavriushenko, Mirka Saarela and Tommi Kärkkäinen

*Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland
magavriu@student.jyu.fi, mirka.saarela@jyu.fi, tommi.karkkainen@jyu.fi*

Keywords: Academic Advising, Learning Analytics, Robust Clustering.

Abstract: The purpose of academic advising is to help students with developing educational plans that support their academic career and personal goals, and to provide information and guidance on studies. Planning and management of the students' study path is the main joint activity in advising. Based on a study log of passed courses, we propose to use robust, prototype-based clustering to identify a set of actual study path profiles. Such profiles identify groups of students with similar progress of studies, whose analysis and interpretation can be used for better institutional awareness and to support evidence-based academic advising. A model of automated academic advising system utilizing the possibility to determine the study profiles is proposed.

1 INTRODUCTION

The credit-based system is used to characterize the requirements and progress of a student in many learning environments. Availability of personalized support is very important constituent of a learner's success (Nguyen et al., 2008). Academic advising (AA) is an iterative collaboration process between student, academic adviser, and academic institution, to tackle the student retention. Advisers provide versatile assistance to the students during their studies, making the educational experience relevant and supported.

AA activity has a long history dating back to 1870s (Tuttle, 2000). Advising starts, when a student becomes enrolled in higher education, and finishes when the degree has been completed. The purpose of academic advising is to ensure that the students carry out the required studies to graduate. The central activity for this purpose is to support study planning, especially at the beginning of the academic life. Depending on the organizational culture, especially the stability or dynamicity of the course schedule, an academic adviser either needs to ensure that predefined study plan is being followed or that a student knows all the relevant study possibilities. Email, social media, web and wikipages etc. provide means to share the necessary information with the students. However, typically face-to-face discussions take place either regularly (e.g., at the beginning of a semester or academic year) or by students' or advisers' request. For more personalized support, an adviser should know when

a student is in need of a study advice discussion and what is the precise status of the studies.

Preliminary recommendations to do certain, especially compulsory, courses to proceed normally with the major subject studies are provided by departments and advisers. More precisely, for example at the University of Jyväskylä (JYU), all students are required to prepare an electronic personal study plan with the academic adviser from the home department. Advising is organized according to the *satellite model* referring to the distributed responsibility of academic units (Tuttle, 2000). The *engagement model* between advisee and adviser characterize the principal spirit of counselling (Feghali et al., 2011).

In JYU, the study plan and the completed studies create the starting point for a study plan assessment discussion. However, especially in computer science, the actual number of studies that have been made during an academic year are typically less than recommended (Saarela and Kärkkäinen, 2015a). Hence, it is very common that the actual study path deviates from the recommendations and plans, and in such a case an advising intervention is needed. 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 using a learning analytics method (Chatti et al.,

2012), more precisely robust clustering (Äyrämö, 2006; Saarela and Kärkkäinen, 2015a), 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. One can explicitly link an individual student to student peers with 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 one 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 (Saarela and Kärkkäinen, 2015a). By automating general student profiling it is possible to provide essential support for adaptive on-line advising along the lines suggested, e.g., in (Nguyen et al., 2008; Henderson and Goodridge, 2015). Individual student's perspective, from self-regulated learning and study planning point of view without academic advising interface, was thoroughly addressed in (Auvinen, 2015) (see also (Auvinen et al., 2014)).

The contents of this article is as follows: after the introduction, we provide background on academic advising and personalized student support in Section 2. Then, in Section 3, we describe the data and the robust clustering method, and introduce three cases to construct student profiles to support academic advising. In Section 4 features and a model of an Academic Adviser system with automated mechanisms in it are proposed. The work is concluded in Section 5.

2 BACKGROUND

Academic advising 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 (Henderson and Goodridge, 2015). 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. (Noaman and Ahmed, 2015). 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 (Noaman and Ahmed, 2015).

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 Academic Advising

Student voices on AA were raised in only some articles. El-Ansiri et al. (Al-Ansari et al., 2015) 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 (Drozd, 2010) 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 (Drozd, 2010). Dougherty (Dougherty, 2007) 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 (Biletskiy et al., 2009). The authors proposed course outline data extractor application, which helps in recognizing similar or comparable courses between different institutions, also help-

¹<https://www.jyu.fi/itp/en/korppi-guide>

ing 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 (Nguyen et al., 2008) 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.

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 (Henderson and Goodridge, 2015), 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 (Feghali et al., 2011). 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 (Latorre-Navarro and Harris, 2015). 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 (Wen and McGreal, 2015) for AA. This approach helps tackling a dynamic and complex individualized study planning and scheduling problem. As well as in (Al-Sarem, 2015) 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 (Al-Ansari et al., 2015).

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 (Huang et al., 2007). In the study (Chen et al., 2005) 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 (Huang et al., 2007) 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 (Chen, 2009a), 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 (Hwang et al., 2010). 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 (Werghi and Kamoun, 2009) 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 (Idris et al., 2009) using neural networks, self organizing maps and the back-propagation algorithm.

A very closely related work to ours was reported in (Sandvig and Burke, 2005). 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 (Mostafa et al., 2014) for developing a recommendation system for a suitable major to students based on comparison of the student information and similar historical cases.

3 CREATION OF STUDY PATH PROFILES USING ROBUST CLUSTERING

3.1 Data

To illustrate the proposed approach, we utilized real study records of the Bachelor (BSc) and Master (MSc) students majoring in Mathematical Information Technology (which is comparable to a major in Computer Science at other universities) at the University of Jyväskylä (JYU/MIT). IT administration at the University has recently created a data warehouse of passed courses by all the students, which can be utilized by the departments. On the other hand, the electronic study plan system does not provide direct interface for larger student groups, so both from accessibility and evidence-basedness points of view, we focus on analyzing the real study log of the passed courses. The log was anonymized, keeping student IDs as keys, covering the four calendar years 2012–2015. Note that students can start their studies in the beginning of September (autumn term) or January (spring term). Hence, the original study registry log included a heterogeneous set of BSc and MSc 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. There were 942 male students (81%), with mean amount of studies made 59.9 ECTS. Only 221 female students (19%) were identified, with mean amount of studies made 57.0 ECTS. Hence, most of the students in the log were either in the beginning of their studies or progressing very passively and slowly.

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 MSc thesis, 30 ECTS. Teaching in JYU is organized for four periods during one academic year (plus the sum-

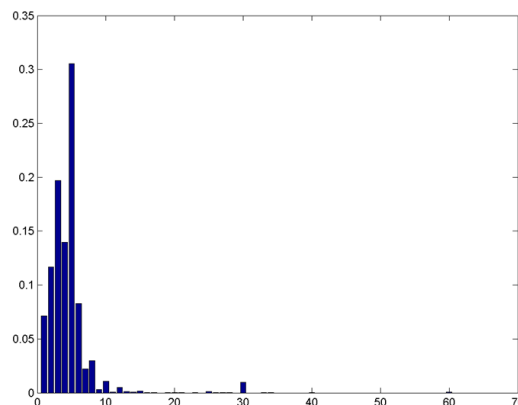


Figure 1: Probability of size of a course.

mer semester) in such a way that a course of ca. 5 ECTS can fit to one period. We conclude that because the passed courses represent both major and minor subject studies, division of the overall learning objects as courses is not optimal. This observation is the first example on how summarization of study log data provides visibility and feedback to the organization. During the course of writing this article, we also found out that the instructions of JYU for preparing the next curriculum for 2017–2019 include strong recommendation to decrease the number of courses with only a few credits.

Next we aggregated how many credits per semester each student had made. Similarly to (Saarela and Kärkkäinen, 2015a), each calendar year was divided into two semesters: the spring term (from January to June) and the autumn term (from July to December). However, since usually only a few courses are completed during the summer (this is illustrated in (Saarela and Kärkkäinen, 2015a)), it was reasonable to divide the calendar year into only two parts for further analysis.

In what follows, we profile, analyse 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 (Saarela and Kärkkäinen, 2015a). 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 (Zaki and Meira Jr, 2014)) 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* (Jain, 2010), also in learning analytics studies (Saarela and Kärkkäinen, 2017). 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 (Huber, 2011) as well as missing values, which can be characterized as special types of outliers. Also for a non-symmetric (skewed) distribution, the sample mean is not necessarily the most efficient estimator and other location estimates might be preferable (Sprent and Smeeton, 2016). Moreover, as explained in (Saarela and Kärkkäinen, 2015a), 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 (Sprent and Smeeton, 2016). Mathematically, the spatial median is the Weber point that minimizes the (nonsquared) sum of the Euclidean distances to a group of given points $\{\mathbf{x}_i\}, i = 1, 2, \dots, n$:

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

Although the basic concept is easily understood and has been extensively discussed in the literature (albeit under various names, see (Drezner and Hamacher, 2001)), its computation is known to be difficult.

In (Äyrämö, 2006), the difficulty of computing the spatial median during partitional clustering was solved with the SOR (Sequential Overrelaxation) algorithm (see (Äyrämö, 2006) for details). Moreover, in the implementation of the resulting *k-spatial-medians* clustering algorithm, only the available (i.e. 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 (Saarela and Kärkkäinen, 2014; Saarela and Kärkkäinen, 2015b; Saarela and Kärkkäinen, 2015). The initialization of the robust clustering method was realized similarly as in (Saarela and Kärkkäinen, 2015): 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.

3.3 Clustered Student Profiles

Similarly to the earlier work in (Saarela and Kärkkäinen, 2014; Saarela and Kärkkäinen, 2015b; Saarela and Kärkkäinen, 2015a; Wallden, 2016), 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 (Sprent and Smeeton, 2016). 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 (Study Amanuensis, 2016).

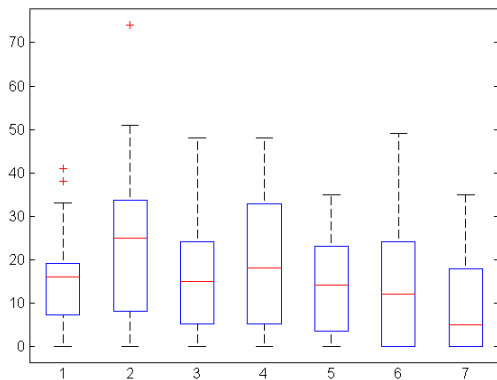


Figure 2: Boxplot for A2012.

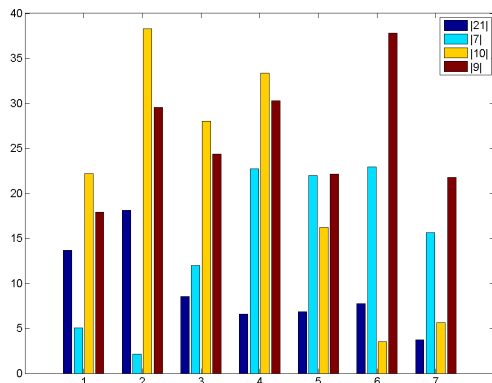


Figure 3: Credit accumulation prototypes for A2012.

A2012

The boxplot in Figure 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 Figure 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 (Study Amanuensis, 2016), by a closer look on the 8 students from the cluster closest to the cen-

troid, these are all older BSc and MSc 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 (Kaihlaivirta et al., 2015) from the same context (department) than here.

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 characteristic 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 characteristic students were again all males, one MSc student and two BSc students. Note that similar finding on the importance of active start in an individual course level was given in (Saarela and Kärkkäinen, 2015a).

Students who are mostly in need of academic advising 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 Figure 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 Figure 4 supports such finding, especially showing smaller variability in the obtained credits between the autumn and spring terms compared to A2012 in Figure 2.

Student group in the need of intrusive academic advising 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

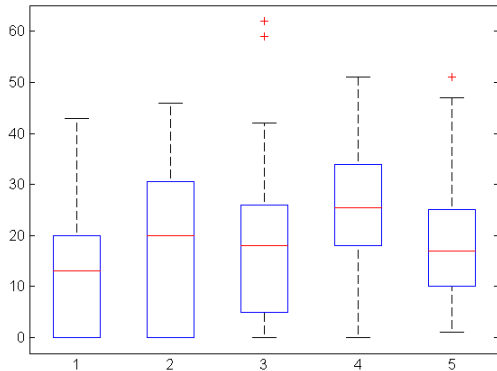


Figure 4: Boxplot for A2013.

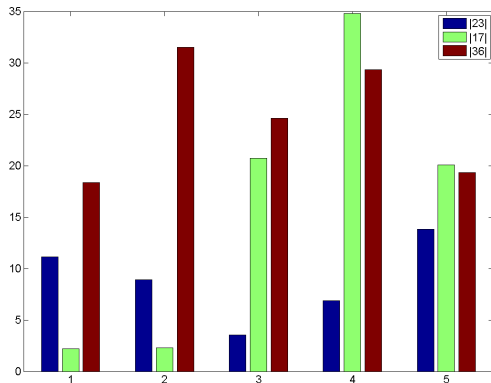


Figure 5: Credit accumulation prototypes for A2013.

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 (Study Amanuensis, 2016), especially the younger students showed signs of low self-regulation during academic advising sessions.

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 Figure 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 calendar 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 academic advising. 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 academic advising. 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 academic advising have improved in the educational organization under study.

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 (Huang et al., 2007).

L. K. Henderson and W. Goodridge (Henderson and Goodridge, 2015) proposed the following division of the *computer supported AA systems* (i.e., Academic Advisers): Basic Computerized Systems, Advanced Automated Systems, and Intelligent Advanced Automated Systems. In this paper, we mostly focus on Advanced Automated and Intelligent Advanced Automated Systems, because these kind of

systems can significantly reduce the human adviser's work and the student load in supporting the study process.

As can be concluded from Section 2.1, 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 (Idris et al., 2009; Chen, 2009a; Leung et al., 2010; Nguyen et al., 2008; Biletskiy et al., 2009; Henderson and Goodridge, 2015).

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 (Minaei-Bidgoli, 2004). 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 (Liu et al., 2008). 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 (Schiaffino et al., 2008; Chen, 2008; Chen et al., 2008; Chen, 2009b; Chen et al., 2005).

As reviewed in Section 2.1, 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 (Fung and Yeung, 2000; Lee, 2001; Brusilovsky, 1998; Lee, 2001; Papanikolaou et al., 2002; Tang and McCalla, 2005). 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 academic advising. 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 (Leung et al., 2010).

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).

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 Figure 6. 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 academic advising 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.

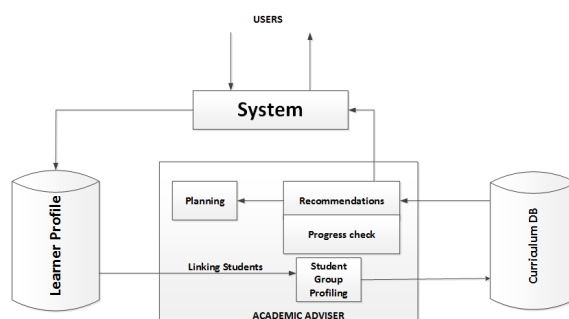


Figure 6: Architecture of the Academic Adviser.

2. Collection of information about the courses and completed studies.
3. Creation of study progress profiles along the lines of Section 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.
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 Academic Advising, related to the system's architecture and main functionality as described above, is presented in Figure 7. 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 Figure 8. 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 Figure 7 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 in them is stored in the database and is automatically changed/refreshed after each passed course. After

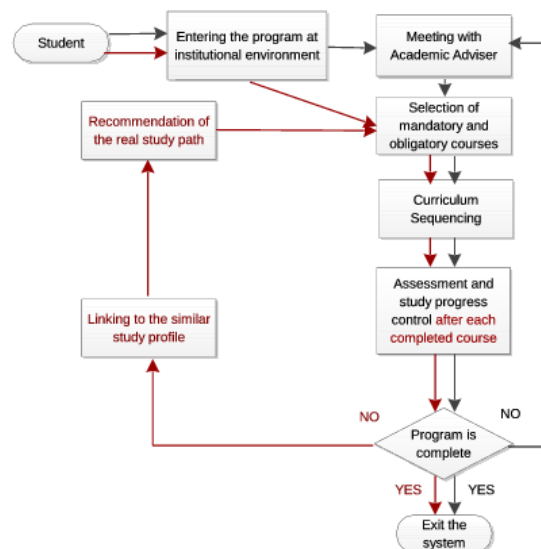


Figure 8: Comparison of the manual and automated processes of learner's study life cycle.

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 academic advising 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.

5 CONCLUSIONS

Academic Advising is an essential part of daily activities in an educational institution and an important component in the learner's study life. Nowadays, we need to be able to create and manage personalized study plans and study paths taking into account learners abilities and regulations of the learning environment. And in order to better help students, Academic Adviser should be able to manage a rich set of information, e.g., on short-range program planning, evaluation of students, and generation of the proper teach-

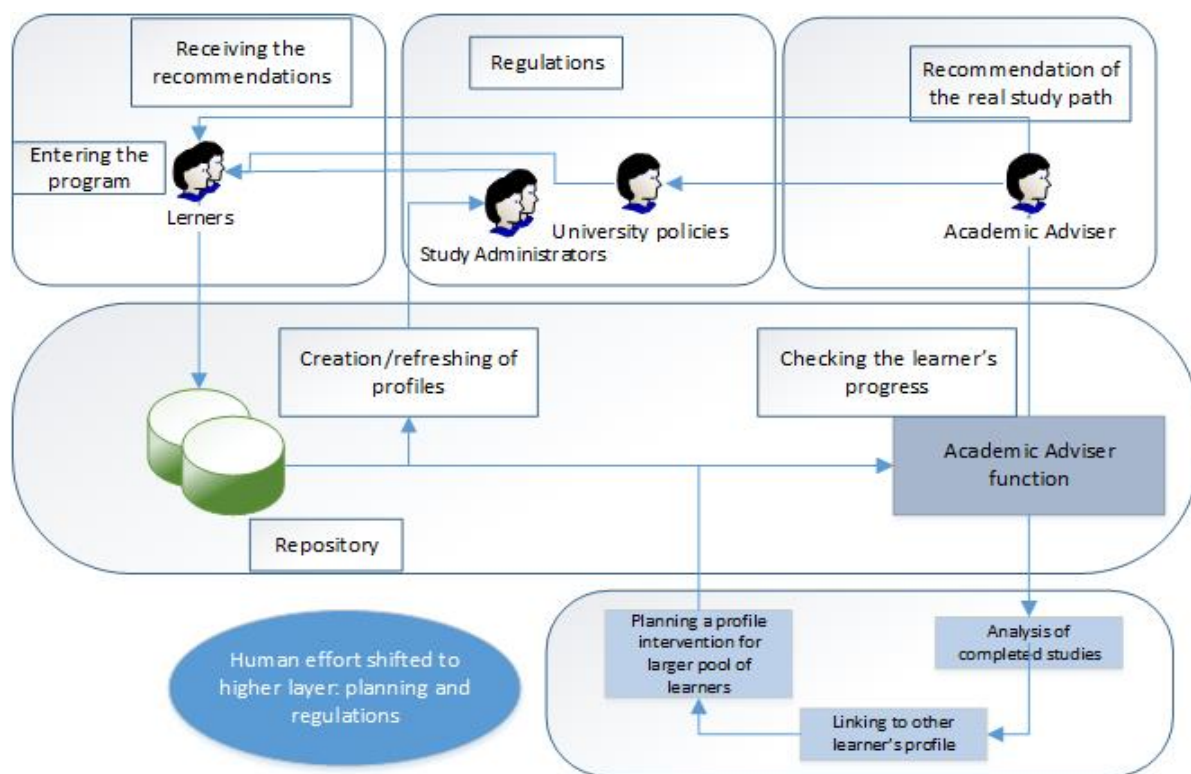


Figure 7: Automated Academic Advising process.

ing schedule, as well as plan possible interventions adaptively for a group of students instead of following all individual students separately. 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 (Al-Ansari et al., 2015).

In this paper, we presented a compact literature review about Academic Advising, mostly focusing on Automated Academic Advising and Intelligent Academic Advising. It was then described how, by using a robust variant of prototype-based clustering method, which is especially suitable for data with missing values, one can create prototypical student group profiles characterizing the overall progress of the studies. This allows academic advisers to provide evidence-based information on the study paths that were actually realized by individual students. Moreover, academic institutions can focus on management and updates on course schedule having an effect on clearly characterized and recognized 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 (Saarela and Kärkkäinen, 2015b).

Then a reference model for automated Academic Advising system was proposed. 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 proposed 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 academic advising.

By continuing the development of the line of work, we could consider the study paths with higher granularity than per semester. Also, the main functionality of the proposed system – to provide 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 – should be properly evaluated. Moreover, better availability of learner's personal information concerning the study entrance criteria and current life situation, e.g., a part-time job or living far from the institute, could support both interpretation of the generated student profiles and better preparation and management of the intervention patterns of academic advising.

REFERENCES

- Al-Ansari, A., El Tantawi, M., AbdelSalam, M., and Al-Harbi, F. (2015). Academic advising and student support: Help-seeking behaviors among Saudi dental undergraduate students. *The Saudi dental journal*, 27(2):57–62.
- Al-Sarem, M. (2015). Building a decision tree model for academic advising affairs based on the algorithm C 4-5. *arXiv preprint arXiv:1511.04026*.
- Auvinen, T. (2015). *Educational Technologies for Supporting Self-Regulated Learning in Online Learning Environments*. Aalto University.
- Auvinen, T., Paavola, J., and Hartikainen, J. (2014). STOPS: a graph-based study planning and curriculum development tool. In *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, pages 25–34. ACM.
- Äyrämö, S. (2006). *Knowledge Mining Using Robust Clustering*, volume 63 of *Jyväskylä Studies in Computing*. University of Jyväskylä.
- Biletskiy, Y., Brown, J. A., and Ranganathan, G. (2009). Information extraction from syllabi for academic e-Advising. *Expert Systems with Applications*, 36(3):4508–4516.
- Brusilovsky, P. (1998). 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'98), San Antonio, TX*.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.
- Chen, C.-M. (2008). Intelligent web-based learning system with personalized learning path guidance. *Computers & Education*, 51(2):787–814.
- Chen, C.-M. (2009a). Ontology-based concept map for planning a personalised learning path. *British Journal of Educational Technology*, 40(6):1028–1058.
- Chen, C.-M. (2009b). Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Systems with Applications*, 36(5):8816–8829.
- Chen, C.-M. et al. (2008). Ontology-based concept map for planning personalized learning path. In *IEEE International Conferences on Cybernetics & Intelligent Systems*.
- Chen, C.-M., Lee, H.-M., and Chen, Y.-H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, 44(3):237–255.
- Dougherty, S. B. (2007). Academic advising for high-achieving college students. *Higher Education in Review*, 4:63–82.
- Drezner, Z. and Hamacher, H. W. (2001). *Facility location: applications and theory*. Springer Science & Business Media.
- Drozdz, D. S. (2010). *Student preferences for academic advisors as transformational leaders*. PhD thesis, Texas A&M University.
- Feghali, T., Zbib, I., and Hallal, S. (2011). A web-based decision support tool for academic advising. *Educational Technology & Society*, 14(1):82–94.
- Fung, A. and Yeung, J. (2000). An object model for a web-based adaptive educational system. In *Proceedings of the IFIP International Conference on Educational Use of Technologies (ICEUT2000)*.
- Henderson, L. K. and Goodridge, W. (2015). AdviseMe: An intelligent web-based application for academic advising. *International Journal of Advanced Computer Science and Applications*, 6(8):233–243.
- Huang, M.-J., Huang, H.-S., and Chen, M.-Y. (2007). Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Expert Systems with Applications*, 33(3):551–564.
- Huber, P. J. (2011). *Robust statistics*. Springer.
- Hwang, G.-J., Kuo, F.-R., Yin, P.-Y., and Chuang, K.-H. (2010). A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning. *Computers & Education*, 54(2):404–415.
- Idris, N., Yusof, N., Saad, P., et al. (2009). Adaptive course sequencing for personalization of learning path using neural network. *Int. J. Advance. Soft Comput. Appl*, 1(1):49–61.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666.
- Kaihlaivirta, A., Isomöttönen, V., and Kärkkäinen, T. (2015). A self-ethnographic investigation of continuing education program in engineering arising from economic structural change. *Studies in Continuing Education*, 37(1):99–114.
- Latorre-Navarro, E. M. and Harris, J. G. (2015). An intelligent natural language conversational system for academic advising. *Editorial Preface*, 6(1).
- Lee, M.-G. (2001). Profiling students' adaptation styles in web-based learning. *Computers & Education*, 36(2):121–132.
- Leung, C. M., Tsang, E. Y., Lam, S., and Pang, D. C. (2010). Intelligent counseling system: A 24 x 7 academic advisor. *Educause Quarterly*, 33(4):n4.
- Liu, C.-I., Wu, S., Chang, M., and Heh, J. (2008). Guiding students to do remedial learning in school campus with learning objects' spatial relations. In *16th international conference on computers in education conference*, pages 249–256. Citeseer.
- Minaei-Bidgoli, B. (2004). *Data mining for a web-based educational system*. PhD thesis, Michigan State University.
- Mostafa, L., Oatley, G., Khalifa, N., and Rabie, W. (2014). A case based reasoning system for academic advising in egyptian educational institutions. In *2nd International Conference on Research in Science, Engineering and Technology (ICRSET2014) March*, pages 21–22.
- Nguyen, T. B., Nguyen, D. N., Nguyen, H. S., Tran, H., and Hoang, T. A. D. (2008). An integrated approach for an academic advising system in adaptive credit-based learning environment. *Journal of Science, Natural Sciences and Technology*, (24):110–121.

- Noaman, A. Y. and Ahmed, F. F. (2015). A new framework for e academic advising. *Procedia Computer Science*, 65:358–367.
- Papanikolaou, K. A., Grigoriadou, M., Magoulas, G. D., and Kornilakis, H. (2002). Towards new forms of knowledge communication: the adaptive dimension of a web-based learning environment. *Computers & Education*, 39(4):333–360.
- Saarela, M. and Kärkkäinen, T. (2014). Discovering Gender-Specific Knowledge from Finnish Basic Education using PISA Scale Indices. In *Proceedings of the 7th International Conference on Educational Data Mining*, pages 60–68.
- Saarela, M. and Kärkkäinen, T. (2015a). Analysing Student Performance using Sparse Data of Core Bachelor Courses. *JEDM-Journal of Educational Data Mining*, 7(1):3–32.
- Saarela, M. and Kärkkäinen, T. (2015b). 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*, pages 156–163.
- Saarela, M. and Kärkkäinen, T. (2015). Weighted Clustering of Sparse Educational Data. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015)*, pages 337–342.
- Saarela, M. and Kärkkäinen, T. (2017). Knowledge Discovery from the Programme for International Student Assessment. In Peña-Ayala, A., editor, *Learning Analytics: Fundamentals, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, pages 229–267. Springer International Publishing, Cham.
- Sandvig, J. and Burke, R. (2005). Aacorn: A cbr recommender for academic advising. Technical report, Technical Report TR05-015, DePaul University.
- Schiaffino, S., Garcia, P., and Amandi, A. (2008). eteacher: Providing personalized assistance to e-learning students. *Computers & Education*, 51(4):1744–1754.
- Sprent, P. and Smeeton, N. C. (2016). *Applied nonparametric statistical methods*. CRC Press.
- Study Amanuensis (November 2, 2016). private communication.
- Tang, T. Y. and McCalla, G. (2005). Smart recommendation for an evolving e-learning system: Architecture and experiment. *International Journal on elearning*, 4(1):105.
- Tuttle, K. N. (2000). Academic advising. *New directions for higher education*, 2000(111):15–24.
- Wallden, L. J. (2016). *Kansainvälisten koulutusarvioiden vertailu koulutuksellisen tiedonlouhinnan keinoin*. Jyväskylä Studies in Computing. University of Jyväskylä.
- Wen, F. L. S. L. D. and McGreal, F. Z. K. R. (2015). e-Advisor: A multi-agent system for academic advising. *International Journal of Advanced Computer Science and Applications*, 6(8).
- Werghi, N. and Kamoun, F. K. (2009). A decision-tree-based system for student academic advising and planning in information systems programmes. *International Journal of Business Information Systems*, 5(1):1–18.
- Zaki, M. J. and Meira Jr, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.