Course Recommender Systems with Statistical Confidence

Zachary Warnes
Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht 6200MD, The Netherlands
z.warnes@student.maastrichtuniversity.nl

Evgueni Smirnov
Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht 6200MD, The Netherlands
smirnov@maastrichtuniversity.nl

ABSTRACT
Selecting courses in an open-curriculum education program is a difficult task for students and academic advisors. Course recommendation systems nowadays can be used to reduce the complexity of this task. To control the recommendation error, we argue that course recommendations need to be provided together with statistical confidence. The latter can be used for computing a statistically valid set of recommended courses that contains courses a student is likely to take with a probability of at least $1 - \epsilon$ for a user-specified significance level $\epsilon$. For that purpose, we introduce a generic algorithm for course recommendation based on the conformal prediction framework. The algorithm is used for implementing two conformal course recommender systems. Through experimentation, we show that these systems accurately suggest courses to students while maintaining statistically valid sets of courses recommended.

Keywords
Recommender Systems, Course Recommendation, Conformal Prediction

1. INTRODUCTION
Recommender systems are systems capable of predicting the preferences of users over sets of items [1]. They can be found almost everywhere in the digital space, shaping the choices we make, the products we buy, the books we read, or the movies we watch. The range of applications of recommender systems has been broadened recently to the education domain, especially in higher education [5]. There are systems reported that provide recommendations for academic choices, learning activities, learning resources, and learning collaborations [14].

Among the recommender systems for academic choices, there exists a particular interest in systems that recommend courses [3]. There is a wide range of such systems that differ in the underlying recommendation mechanism, accuracy, type of recommendations (courses, course sequences, course concentrations), and type of representation. It has been recently recognized that course recommender systems need to be safe [11]; i.e., course recommendations need to be provided with confidence information that will help a student to make a better course selection. There exist different approaches to delivering such confidence information from course preference ranks estimated by the underlying recommendation mechanisms [3, 6, 10, 12] to separate warning modules [11]. The characteristic feature of these approaches is that they are heuristic, and thus they do not provide any theoretical guarantees for the quality of course recommendation.

In this paper, we argue that course recommendations need to be supported with statistical confidence. This confidence will allow computing a statistically valid set of recommended courses that contains courses a student is likely to take with a probability of at least $1 - \epsilon$ for a user-specified significance level $\epsilon$. To achieve this, we employ the well-known conformal-prediction framework [4, 15, 16]. We design a generic algorithm for conformal course recommendation capable of computing statistically valid sets of courses for students. The algorithm is used for implementing two conformal course recommender systems that employ a content-based recommendation mechanism. The first system is instance-based, and the second system is an exemplar-based system [13].

The conformal course recommender systems have been implemented for academic advising of University College Maastricht, a Liberal Arts Bachelor study with an open curriculum. In this study, students personalize their program by selecting courses that align with their academic and personal interests. In total, students choose around 40 out of the 160 possible educational modules; i.e., they create a program by selecting one path out of $\binom{160}{40}$ possible. Our recommender systems are tested to facilitate this process. The initial experimental results show that the systems accurately recommend courses while providing statistically valid sets of courses recommended.

The rest of the paper is organized as follows. The related work is provided in Section 2. Section 3 formalizes the task of course recommendation. The course and student topic models used for course recommendation are briefly described in Section 4. Section 5 introduces the generic algorithm for conformal course recommendation and its instantiations: the instance-based and exemplar-based recommender sys-
tems. The student-course data is described in Section 6. Section 7 provides the experiments and discussion. Finally, Section 8 concludes the paper.

2. RELATED WORK
Course recommender systems received significant attention since the very first publications [12, 18, 17]. Meanwhile, these systems have become very diverse. Following the main trends in recommender-system research there are different types of course recommender systems: content-based systems [10, 11], collaborative-filtering systems [3], hybrid systems [3, 6], and popularity-based ranking systems [6]. Most of these systems are capable of providing (explicitly/implicitly) confidence information for course recommendations. However, this does not give any guarantee for the quality of course recommendation in a statistical sense.

Confidence-based recommender systems have been proposed based on collaborative filtering. The first system is based on group recommender systems [8], and the second one is based on matrix factorization [7]. Both systems can be directly applied for course recommendations, however, under assumptions typical for collaborative filtering. For example, plenty of data is available; the course order does not matter. In this context, we note that we propose a generic algorithm for conformal course recommendation that is not tailored to the recommendation mechanism: collaborative filtering or content-based filtering. The only requirement to apply this algorithm is to have a function that estimates the typicality of a course w.r.t. other courses taken by a student (see conformity functions in Section 5).

3. RECOMMENDATION TASK WITH CONFIDENCE
Let $T$ be a set of topics $t$ considered in a set $C$ courses $c$. To indicate the degree of presence of topic $t$ in course $c \in C$ we employ weight $w_{c,t}$. Topic weights $w_{c,t}$ of course $c \in C$ represents a topic model of this course. We assume that the topic models (i.e. topics’ weights $w_{c,t}$) are provided initially for all the courses $c \in C$. We describe our approach to derive these models in the next Section.

The courses $c \in C$ are given for a set $S$ of students $s$. To indicate the degree of student $s \in S$ masters topic $t$ we employ weight $w_{s,t}$. Topic weights $w_{s,t}$ of student $s \in S$ represents a topic model of the student w.r.t. courses $c \in C$. Thus, they are computed w.r.t. set $C_s$ of courses student $s$ has taken; i.e. for any topic $t \in T$ we have:

$$w_{s,t} = \frac{\sum_{c \in C_s} w_{c,t}}{|C_s|},$$  

(1)

If we assume a specific ordering of the topics $t \in T$, then:

- the topics’ weights $w_{c,t}$ for course $c$ form a topic-model vector $w_{c}$ for $c$, and
- the topics’ weights $w_{s,t}$ for student $s$ form a topic-model vector $w_{s}$ for $s$.

The topic-model vectors $w_c$ and $w_s$ “live” in the same space $W$. Due to the number $|T|$ of all the topics, we employ the cosine similarity over $W$. It can be used to compute similarity for any two topic-model vectors that represent courses and students.

The topic-model vectors $w_c$ of all the courses $c \in C$ form the course data set $W_C$ defined as $\{w_c\}_{c \in C}$. Analogously, the topic-model vector $w_s$ of all the students $s \in T$ form the student data set $W_S$ defined as $\{w_s\}_{s \in T}$. In this context, we introduce the recommendation task considered in this paper. Given a course data set $W_C$, a student data set $W_S$, and a student $s \in S$ with topic-model vector $w_s \in W_S$, the task is to compute a recommendation set $C_s^+ \subseteq C \setminus C_s$ that contains courses that indeed fit student $s$ with a probability at least $1 - \epsilon$ for a predefined significance level $\epsilon \in [0, 1]$.

4. COURSE AND STUDENT TOPIC MODELING
We employed the topic-modeling approach proposed in [11]. The set $T$ of topics $t$ was identified from the course descriptions using the Latent Dirichlet Allocation (LDA) generative model [2]. Each topic $t \in T$ is given by a probability distribution over the vocabulary derived from all the descriptions. Thus, each course $c \in C$ is represented by topics $t$, which words are present in the description of that course. Student topic models are derived based on the topics courses using formula (1).

5. CONFORMAL COURSE RECOMMENDATION
This section introduces a conformal course recommendation. First, we present a conformal test for course inclusion and a generic algorithm for conformal course recommender systems. Then we provide two instantiations of this algorithm.

5.1 Generic Conformal Course Recommender
Consider a particular student $s \in S$ with her set $C_s$ of courses. We assume that student $s$ is represented by a probability distribution $P_s$; i.e. $P_s$ has generated the course set $C_s$ for $s$. Thus, to decide whether to recommend a new course $c \notin C_s$ for student $s$, we perform a statistical test of the null hypothesis that the set $C_s \cup \{c\}$ is generated by the student distribution $P_s$ under the exchangeability assumption [15] 1.

We implement the statistical test according to the conformal-prediction framework [15]. It makes use of course conformity scores. The conformity score $\alpha_c$ of a course $c$ is defined as a score that indicates how typical $c$ in set $C_s \cup \{c\}$. The conformity score $\alpha_c$ is computed by a course conformity function $A$. The latter is a mapping from $2^C \times C$ to $\mathbb{R} \cup \{+\infty\}$; i.e. it returns for any course set $C_s$ and any course $c$ a score $\alpha_c$ that indicates how typical is course $c$ for the courses in $C_s \cup \{c\}$. Depending on the implementation of the conformity function for the course and student topic models, we can have recommender systems based on content/collaborative filtering (See the next section).

The conformity score $\alpha_c$ of a new course $c$ is used as a test statistic for the null hypothesis that the set $C_s \cup \{c\}$ is generated by the student distribution $P_s$ according to the

1We note that the exchangeability assumption is weaker than the well-know i.i.d. assumption.
Thus, to guarantee valid recommendation sets $C^s \subseteq C \setminus C_s$ that contains courses that fit students with a probability at least $1 - \epsilon$ we need to set the course significance level $\epsilon_c$ according to formula (3) when we initialize the generic conformal course recommender algorithm from Algorithm 1.

\[ \epsilon_c = \frac{t}{n} \epsilon \]

3.2 Content-based Conformal Course Recommender Systems

The generic conformal course recommender algorithm can be instantiated if we specify the course conformity function $A$. This function can be done using different recommender mechanisms, e.g., collaborative filtering or content-based filtering. In this paper, we assume the existence of topic model
vectors of courses and students that fit the content-based filtering scenario (see Section 3). That is why we propose conformity functions for two content-based conformal course recommender systems specified below.

The first system is an instance-based conformal course recommender system (ICCRS). Any student \( s \in S \) is represented by a set of topic-model vectors (instances) \( w_c \) of the courses \( c \in C_s \) she has taken. In this context the course conformity function \( A \) outputs for any course \( c \in C \) and course set \( C_s \) of student \( s \in S \) an averaged similarity of \( c \) with courses in \( C \); i.e. 
\[
\frac{1}{|C_s\setminus\{c\}|} \sum_{c' \in C_s \setminus \{c\}} \cos(w_c, w_{c'})
\]
where \( \cos \) is the cosine similarity.

The second system is an exemplar-based conformal course recommender system (ECCRS). It employs topic-model vectors \( w_s \) (exemplar) of student \( s \in S \) computed using formula (1). In this context the course conformity function \( A \) outputs for any course \( c \in C \) and course set \( C_s \) of student \( s \in S \) a value equal to \( \cos(w_c, w_s) \), where topic-model vector \( w_s \) of student \( s \in S \) is based on the courses in \( C_s \setminus \{c\} \) and \( \cos \) is the cosine similarity.

The computational complexity of ECCRS is higher than that of ICCRS since, for any student, we need to recompute her topic-model vectors \( w_s \) by excluding courses one by one. However, ECCRS has better explanation capabilities. The topic-model vector \( w_s \) of student \( s \) represents the current levels of topic mastering, and the topic-model vector \( w_c \), of course, \( c \) represents the topics covered in the course. Thus, the cosine match can explain why the course has been selected/rejected.

6. STUDENT-COURSE DATA

ICCRS and ECCRS have been implemented as course recommender systems for University College Maastricht. The college has provided course enrollment data from 2008 to 2017. This data includes course and student identifiers, grades for each course, details regarding course assessment, ECT credits, and course descriptions. The course descriptions facilitate the construction of topic values for both the student model and the course model. The calculation of topic values is with LDA, and an optimal number of topics is determined through maximum likelihood estimation. This optimization results in sixty-five topic areas representing the course catalog [11]. We remove modules without descriptions from consideration. In total, 143 courses and 2422 students enrolled in at least one course remain.

The rates of course enrollments vary widely between each course. Registration in the majority of courses offered occurs only a few times over the entire period, see in Figure 1. The modules provided are updated each year, reflecting the changes to the course catalog via dropping courses and course code changes. Several introductory courses, required courses, and projects make up a significant portion of all enrollments. Most students at UCM need eighteen periods to complete their education. Nevertheless, some students enroll in over twenty periods. See Figure 2. Each recommender system focuses on a subset of twelve periods representing two years at UCM. The subset is refined further by selecting only students starting in the fall intake semester. These restrictions increase the standardization of students for our systems, and balance for the diversity of enrollment patterns present in an open-course curriculum. Our recommender systems use the remaining 1018 students that fall within these boundaries.

7. EXPERIMENTS
This section presents experiments of ICCRS and ECCRS on the student-course data provided by University College Maastricht (see Section 6). First, an experimental setup is given, followed by results and discussion.

7.1 Setup
We validate ICCRS and ECCRS on the student-course data in the order of study periods. Assume that we have \( M \) number of periods \( P_1, \ldots, P_M \) in which a student studies towards her degree (for our data \( M = 18 \)). We denote by \( C_s(P_m) \subseteq C_s \) the set of courses that student \( s \) has taken in period \( P_m \) for \( m < M \). Given new period \( P_{m+1} \) together with the set \( \bigcup_{s=1}^{m+1} C_s(P_m) \) of courses \( s \) has taken in any period, we test our recommender systems by checking whether the recommended sets \( C^*_s \) of courses for \( P_{m+1} \) includes the courses \( C_s(P_{m+1}) \) that student \( s \) indeed has taken in \( P_{m+1} \).

For the validation process, we estimate the average error of \( \epsilon \), and the average size of \( SR \) of the recommended sets \( C^*_s \) of courses. We then use these statistics to study ICCRS and ECCRS as conformal predictors and as recommender systems.

In the first study, when we investigate ICCRS and ECCRS as conformal predictors, we are interested in establishing the validity and informational efficiency of the systems (check Sub-section 5.1). In the second study, when we investigate ICCRS and ECCRS as recommender systems, we are interested in estimating the error of the systems over the periods when we employ the recommended sets \( C^*_s \) on a given course significance level \( \epsilon \). In our experiments, we use course significance levels \( \epsilon \) of 0.05 and 0.1.

7.2 Results and Discussion
Figures 6 and 7 present the error plots and size plots of the recommended sets \( C^*_s \) of ICCRS and ECCRS, respectively, for course significance level \( \epsilon = 0.05 \). The error curves are very close to the diagonal \((0,0) - (1,1)\), which means the error is close to the course significance level \( \epsilon \). For ICCRS, the error is bounded mainly from above. For ECCRS, the error is bounded mainly from below. This bounding indicates that the systems are valid given sufficient information, especially ICCRS, which is conservatively valid [15].

The conservative validity of ICCRS explains why the averaged size \( SR \) of the recommended sets \( C^*_s \) is higher than that of ECCRS. Thus, we may conclude that the informational efficiency of ECCRS is better in our experiments.

![Figure 5: Period errors of ICCRS and ECCRS on course significance level \( \epsilon \) of 0.05 and 0.1](image)

For the validation process, we estimate the average error of \( \epsilon \), and the average size of \( SR \) of the recommended sets \( C^*_s \) of courses. We then use these statistics to study ICCRS and ECCRS as conformal predictors and as recommender systems.

In the first study, when we investigate ICCRS and ECCRS as conformal predictors, we are interested in establishing the validity and informational efficiency of the systems (check Sub-section 5.1). In the second study, when we investigate ICCRS and ECCRS as recommender systems, we are interested in estimating the error of the systems over the periods when we employ the recommended sets \( C^*_s \) on a given course significance level \( \epsilon \). In our experiments, we use course significance levels \( \epsilon \) of 0.05 and 0.1.

7.2 Results and Discussion
Figures 6 and 7 present the error plots and size plots of the recommended sets \( C^*_s \) of ICCRS and ECCRS, respectively, for course significance level \( \epsilon = 0.05 \). The error curves are very close to the diagonal \((0,0) - (1,1)\), which means the error is close to the course significance level \( \epsilon \). For ICCRS, the error is bounded mainly from above. For ECCRS, the error is bounded mainly from below. This bounding indicates that the systems are valid given sufficient information, especially ICCRS, which is conservatively valid [15].

The conservative validity of ICCRS explains why the averaged size \( SR \) of the recommended sets \( C^*_s \) is higher than that of ECCRS. Thus, we may conclude that the informational efficiency of ECCRS is better in our experiments.

- ICCRS and ECCRS produce accurate recommended sets of courses with an acceptable error.
- ICCRS is more accurate than ECCRS. This difference can be explained by the fact that ICCRS is more conservatively valid.
- the course significance level \( \epsilon \) plays a substantial role: for 0.05, the error of recommended sets \( C^*_s \) is much lower. However, this comes with a price: the size of the recommended sets is bigger when epsilon is lower.

8. CONCLUSION
This paper shows that safe course selection can be obtained if recommendations are supported with statistical confidence. The statistical confidence can be used for computing a statistically valid set of recommended courses that contains courses a student is likely to take with a probability of at least \( 1 - \epsilon \) for a user-defined significance level \( \epsilon \).
We have developed a generic conformal course recommender algorithm that outputs recommendations supported by statistical confidence. The algorithm has been instantiated in the form of two confidence-based course recommendation systems. The systems are essentially content-based: the first is an instance-based recommender system with relatively high accuracy. The second system is an exemplar-based system with a lower accuracy but with better explanatory capabilities. The experiments showed that both systems accurately suggest courses to students while providing statistically valid sets of courses recommended.
REFERENCES