




A comparative analysis of different recommender systems for university major and career domain guidance

Christine Lahoud¹ · Sherin Moussa^{1,2}  · Charbel Obeid³ · Hicham El Khoury⁴ · Pierre-Antoine Champin³

Received: 1 August 2022 / Accepted: 14 December 2022 / Published online: 28 December 2022
© The Author(s) 2022

Abstract

Academic advising is inhibited at most of the high schools to help students identify appropriate academic pathways. The choice of a career domain is significantly influenced by the complexity of life and the volatility of the labor market. Thus, high school students feel confused during the shift period from high school to university, especially with the enormous amounts of data available on the Web. In this paper, an extensive comparative study is conducted to investigate five approaches of recommender systems for university study field and career domain guidance. A novel ontology is constructed to include all the needed information for this purpose. The developed approaches considered user-based and item-based collaborative filtering, demographic-based recommendation, knowledge base supported by case-based reasoning, ontology, as well as different hybridizations of them. A case study on Lebanese high school students is analyzed to evaluate the effectiveness and efficiency of the implemented approaches. The experimental results indicate that the knowledge-based hybrid recommender system, combined with the user-based collaborative filtering and braced with case-based reasoning as well as ontology, generated 98% of similar cases, 95% of them are personalized based on the interests of the high school students. The average usefulness feedback and satisfaction level of the students concerning this proposed hybrid approach reached 95% and 92.5% respectively, which could be a solution to similar problems, regardless of the application domain. Besides, the constructed ontology could be reused in other systems in the educational domain.

Keywords Recommender system · Ontology · CBR · University major · Career domain guidance

✉ Sherin Moussa
sherin.moussa@ufe.edu.eg; sherinmoussa@cis.asu.edu.eg

Extended author information available on the last page of the article

1 Introduction

Choosing a university major or a career domain is a challenging task overflowing with concern that makes students distracted (Zhang, 2021). A university major is the study field at the university, such as computer science, architecture, applied languages, management, arts, law, etc. Furthermore, students do not have enough knowledge about the existing career domains, and with the enormous amounts of data on the Web, they do not find easily the needed information (Chamandy & Gaudreau, 2019). In most cases, students know the careers in their environment, such as their parents and family, and discover other domains through career fairs or everyday life. This lack of knowledge forces them to choose an academic major that does not necessarily correspond to their expectations and would affect their career domain in life. In many situations, the career decision is difficult to undo due to various reasons, i.e., financial, familial, personal restrictions, etc., which may create a seed of regret and dissatisfaction with people's lives (Budjanovcanin & Woodrow, 2022). Therefore, high school students require guidance to balance their interests with the available universities and majors.

The recommender systems (RSs) have proven their ability to assist users with personalized recommendations in many applications (Aggarwal, 2016; AlBanna et al., 2016). RSs have been utilized as the most significant machine learning tools to predict users' behaviors and recommend personalized items, i.e., courses, articles, books, movies, playlists, products, etc. RSs collect several types of information about the users, including interests, preferences, etc. Two categories of information are usually integrated into RSs in order to suggest adequate recommendations: (1) The characteristic information about the items, such as keywords and categories, and the users, such as preferences, interests, and profiles; (2) The user-item interaction information, such as ratings, reviews, likes, and total of purchases. Every RS employs a filtering technique to retrieve suitable suggestions and items for the users. These techniques are mainly categorized into various types as follows:

- *Collaborative filtering* recommender systems (CF), which are based on model-based (clustering, regression, etc.) or memory based (user-item interactions) techniques.
- *Content-based* recommender systems (CB), which depend on the characteristics of information.
- *Knowledge-based* recommender systems (KB), which adopt either ontology-based, case-based or constraint-based techniques.
- *Demographic-based* recommender systems (DF), which depend on the users' demographic data.
- *Hybrid* recommender systems, which represent a combination of two or more filtering techniques.

In the educational domain, most of the existing RSs recommend courses or materials to learners, while the concern of recommending personalized higher

education studies and career domain guidance to learners has been neglected. Accordingly, this study investigates the appropriate RS that can answer the following questions:

- What is the appropriate RS approach to process high school students' profiles, interests, education, demographic, and career knowledge to provoke personalized recommendations?
- What are the solutions to overcome the limitations of traditional RS approaches?
- Is applying one RS approach sufficient to get accurate results? Or should we apply a hybrid one? In the latter case, what is the suitable hybridization technique for a uniform hybrid RS?

Thus, the main contributions in this study can be summarized as follows:

- This study exclusively investigates different recommender systems for university study field and career domain guidance. Many recommender systems have been conducted in the field of education. To the best of our knowledge, none of them recommended higher education study fields (majors) to high school students based on their profiles.
- Most of the studies recommended educational resources or activities for learners. To the best of our knowledge, this research study is specialized in guiding high school students towards higher education paths by recommending universities, university's majors, and career domains.
- A detailed comparison is conducted by investigating five approaches of RSs to address the problem of university majors and career guidance. These RS approaches are: (1) the stand-alone user-based and item-based CF RSs, (2) the stand-alone DF RSs, (3) the stand-alone KB RSs supported by Case-Based Reasoning (KB using CBR), (4) the stand-alone KB RS supported by ontology and CBR (KB using CBR + Ontology), and (5) the Hybrid RS combined with CBR, ontology, and user-based CF (KB using CF + CBR + Ontology).
- We uniquely propose the fifth approach, the Hybrid RS incorporated with user-based CF, ontology and CBR (KB using CF + CBR + Ontology), to answer the presented research questions.
- This proposed hybridization is domain-independent, as it can be extended to solve similar challenges in other domains by defining the corresponding domain of knowledge.
- We uniquely introduce the GraduateOnto ontology to describe the domain knowledge of graduates' students from universities. Thus, it could be reused in other problems in the educational domain.

The rest of the paper is structured as follows. In the next section, we present the works related to RSs in the educational domain. Section 3 introduces the data collection and preprocessing processes. Section 4 discusses the five investigated approaches to assist high school students. Section 5 provides a thorough discussion of the experimental results of these approaches, while Section 6 concludes the paper and highlights the expected future works.

2 Related recommender systems in education

Several approaches were proposed to develop recommender systems, which provoke recommendations to their users as per certain criteria that meet their preferences (Isinkaye et al., 2015; Jumaa et al., 2017; Maher et al., 2020; Moussa et al., 2020). However, these approaches make the prediction process fits a specific domain and dataset complexity. In this section, we present the main related works of different RS techniques in the education domain, except the content-based technique, because it is not compatible with our data nature. As for the CF technique, many problems have been detected, such as (Breese et al., 2013):

- **Cold start:** The CF technique requires the previous users' history, like the users' ratings as well as activities to explore accurate recommendations. This problem occurs when the used data do not comprise enough interests and ratings. Therefore, reliable recommendations become hard to provide. Usually, the *cold start* issue happens due to three main reasons: a new active user, a new community, or a new item added to the system (Schafer et al., 2007). For instance, if a new active user asks for a recommendation, the system finds difficulty to match him/her to similar users, since minimal history exists about his/her activities or ratings in the database. Hybrid systems are used to overcome this problem.
- **Sparsity:** This concern arises when the matrix table of items and users is broadly scattered, which decreases the accuracy of recommendations, having all the available items without any rating from the previous users in the system. Hybridization is commonly used to improve recommendation techniques and to solve this issue. For example, combining the CF and DF recommendation techniques is one method to minimize the sparsity problem of the CF algorithm.
- **Grey-sheep:** Odd recommendations result in such concern, in which the user might have other variant characteristics that do not meet other users (de Campos et al., 2010). This may happen when a user neither complies nor contradicts with any user. Grey-sheep issue can increase the error rate in recommendations and consequently affects the precision of the RS. Furthermore, this issue possibly would negatively affect the predictions for the rest of the community in the dataset (Bruke, 2002).
- **Scalability:** Enormous groups of users and items exist in several environments in which CF systems make their recommendations. Hence, great computation power is necessary to compute recommendations. Dimensionality reduction and clustering techniques are ways to overcome this challenge.
- **Handling high dimensional data:** Elemental recommender filtering approaches cannot maintain high dimensional data, which encompass many attributes. Minimizing the number of attributes could be one solution to this concern or considering a hybrid RS that can handle huge volumes of data.
- **Handling heterogeneous datatypes:** Elemental recommender filtering approaches cannot maintain heterogeneous datatypes, in which the hybrid RSs can manipulate heterogeneous data.

Thus, the scope of our related works is on the CF technique combined with other techniques in hybrid RSs to overcome these limitations. The following subsections present 3 main categories of RSs in the education field: Ontology-based, CBR-based, and hybrid RSs.

2.1 Ontology-based RSs

An ontology is a formal, precise specification of a shared conceptualization (Guarino et al., 2009). It is formal, as it is written in a formal syntax with semantics allowing it to be understandable and interpretable by machines, while it is explicit because its concepts and the relation between them are explicitly defined. Besides, it is shared since the ontology represents a domain knowledge agreed upon and shared by a group of persons (Guarino et al., 2009). The conceptual model of ontology permits reasoning at all concept levels. Hence, an ontology-based RS is an approach of knowledge-based RS techniques that is very popular in the e-learning domain due to its capability to cluster the learners' models based on their educational background, learning style, study trajectory, and knowledge level (Amane et al., 2022; Tarus et al., 2018). In addition, it resolves the cold-start problem (Jeevamol & Renumol, 2021). Numerous ontology-based RSs have been developed with the association of many different recommendation techniques (Rahayu et al., 2022).

(Romero et al., 2019; Shishehchi et al., 2012) presented an ontology-based system to recommend suitable materials to learners. The used ontology integrated the learners' and learning materials' knowledge. Similarly, (Bouhri & Bahaj, 2019) recommended learning materials based on ontology and Semantic Web Rule Language (SWRL) rules, taking into account the learners' learning context. In (Assami et al., 2019), an ontology-based RS was proposed to recommend personalized Massive Open Online Courses (MOOC) resources to learners according to their pace of learning, cognitive learning style, learners' profile, and learning history. (Capuano et al., 2014), built an adaptive e-learning RS called "IntelligentWebTeacher", combining CF and KB techniques and supported by ontology. Another ontology-based hybrid-filtering system called the ontology-based personalized course recommendation (OPCR) was proposed by (Ibrahim et al., 2019) to recommend a higher education course at the university based on the learner's profile and the course content. It combined the CF, KB and CB techniques. In (Sarwar et al., 2019), CBR, neural networks and ontology were combined to recommend personalized content to learners according to their profiles and context awareness, in order to enhance the degree of learner's productivity. In addition, (Qomariyah & Fajar, 2019) proposed an e-learning RS to recommend material content to learners according to their learning style based on the Active Pairwise Relation Learner (APARELL) logic approach and ontology. Authors in (Gulzar et al., 2018) presented the Personalized Course Recommender System (PCRS) based on a hybrid approach of N-Grams queries and ontology to recommend courses to researchers in order to help them choose the appropriate courses in seminal years for time gain and better research.

2.2 CBR-based RSs

CBR (Perner, 2019) is an artificial intelligence technique applicable to problem-solving and learning where earlier cases are available. It is the process of addressing a new problem based on the solutions of similar prior problems, retrieved from a library of prior cases called case-base. CBR-based RSs are also considered as KB RS. Unlike other RSs, a CBR-based RS does not need to save an enormous volume of data about items rating or specific users. The CBR is a specific information retrieval technique extensively used in nearest-neighbor RSs. Several CBR-based RSs have been proposed in the education domain by developing various recommendations techniques.

(Sandvig & Burke, 2005) proposed the Academic Advisor Course Recommendation Engine (AACORN) that implements CBR based on the knowledge of past cases. It integrated knowledge such as past students' experience and courses' history to guide learners in choosing appropriate courses. (Gil et al., 2012) proposed the Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments (AIREH) that can retrieve and incorporate varied personalized labeled educational content acquired from diverse environments by a CBR system. In (Bousbahi & Chorfi, 2015), a CBR-based RS was proposed to recommend the most suitable MOOCs from different resources in reply to a particular request of the learner based on his/her profile, requirements, and knowledge. Another assistant RS was introduced by (Duque Méndez et al., 2018) to guide learners in choosing educational material based on CBR. In (Salam & Fathurrahmad, 2021), a student final project RS was proposed to improve the quality of final assignments in universities. The system was based on CBR to detect a list of research topics and used programming languages in similar projects. Authors in (Gomez-Albarran & Jimenez-Diaz, 2009) presented a CBR approach for personalized recommendations and learners' authoring tasks in online repositories of Learning Objects (LOs), combining CB filtering with CF mechanisms. The learners' authoring tasks included the integration of ratings of the new as well as existing LOs.

2.3 Hybrid RSs

Hybrid RSs have widely shown improved outcomes rather than any standalone filtering technique. Several hybrid combinations between CF and DF techniques have been proposed. Such hybridization minimizes the limitations of CF, such as the sparsity and the cold start concerns because the DF technique does not need the user's rating history. As presented, most of the related works more concentrate on recommending learning to learners' content using different techniques, rather than helping students to find the academic path corresponding to their interests. The applied techniques in these works are summarized in Fig. 1 to demonstrate the strengths and weaknesses of the used hybridization approaches.

For instance, (Schafer et al., 2007) proposed the hybridization of CF and DF approaches to improve the movie recommendation quality. In (Xia et al., 2009),

<i>Hybridization model</i>	<i>Advantages</i>	<i>Drawbacks</i>
<i>KB + CF Memory-based</i>	<ul style="list-style-type: none"> • Reduced the cold-start problem • Reduced the sparsity issue • The accuracy of the recommendations of this hybridization outperforms the memory-based CF predictions. • Fast replying when user's preferences are modified. 	<ul style="list-style-type: none"> • It is not scalable for large datasets. • It needs knowledge engineering.
<i>KB + CF Model-based</i>	<ul style="list-style-type: none"> • The accuracy of the recommendations of this hybridization outperforms the model-based CF predictions. • It has a scalability feature. • Fast replying when user's preferences are modified. 	<ul style="list-style-type: none"> • It needs knowledge engineering. • The hybridization of the memory-based CF with the KB provided better results than this system.
<i>DF + CF Memory-based</i>	<ul style="list-style-type: none"> • Reduced the cold-start problem. • The accuracy of the recommendations of this hybridization outperforms the memory-based CF predictions. 	<ul style="list-style-type: none"> • It is hard to acquire demographic data. • It is not scalable for large datasets.
<i>DF + CF Model-based</i>	<ul style="list-style-type: none"> • The accuracy of the recommendations of this hybridization outperforms the model-based CF predictions. • It has a scalability feature. 	<ul style="list-style-type: none"> • Find difficulty in obtaining demographic data • The hybridization of the memory-based CF with the DF provided better results than this system.

Fig. 1 Comparison of the hybridization of RS techniques

an augmentation item-based CF hybrid system was presented using demographic data to predict missed data such as age and occupation information. In (Agarwal et al., 2017), the authors used users' demographic data instead of users' rating history to generate accurate movie recommendations and overcome the CF cold-start issue. Eventually, minor hybridization approaches were addressed using three filtering techniques. In (Benouaret, 2017), the demographic, semantic, and CF core techniques were combined to propose a hybridization strategy to reinforce the experience of visitors in tourist places and museums. Each method was adapted to a specific stage of the museum visit. The demographic approach was applied to overcome the CF cold-start problem, the semantic approach provoked recommendations semantically close to their previous appreciated visits, whereas the CF approach recommended visits previously liked by similar users.

In education, hybrid RSs have been used in most cases to recommend learning activities or resources to learners (Deschênes, 2020; Tarus et al., 2017). Indeed, the authors in (Farzan & Brusilovsky, 2006) worked on developing an RS based on an adaptive community to recommend appropriate courses to active learners. They analyzed learners' career goals by implementing a social navigation technique. Protus was presented in (Klašnja-Milićević et al., 2011) as a programming tutoring based on the learners' knowledge levels and interests. In (Chavarriga et al., 2014), a KB with CF technique was introduced in order to advise learning materials, helping learners achieve advanced competence levels using an online course platform, whereas in (Tarus et al., 2017), a KB hybrid RS was proposed to advise e-learning materials to learners based on sequential pattern mining with ontology. Moreover, authors in (Rodríguez et al., 2015) introduced a student-centered LO RS that combined CB, KB, and CF techniques, in which the learner's model/profile was used to

adapt the LOs retrieved from the LO databases, considering the descriptive meta-data stored for the objects. Yet, a hybridization of CF, KB, and DF approaches was not tackled.

3 Data collection and preparation

Since our study focuses on assisting high school students to conveniently decide about their higher education choices, the expected recommendations will be based on the university graduates' educational trajectories. Unfortunately, the required data to construct an adequate knowledge base are not available. Moreover, it is very challenging to be obtained online, where users are unwilling to reveal their data. Thus, we gathered the required data by disseminating an online survey including 55 questions. The survey was created in bilingual form (English and French). The survey's dissemination process included the graduates of the Lebanese university in three governorates: South, North, and Beirut. The online survey was posted on social media for three months and sent by email to many mailing lists. A real-world dataset was collected of 869 university graduate profiles and 20,000 high school course ratings.

In order to evaluate the five investigated hybrid RS approaches to generate recommendations for universities, university majors, and career fields, the following four criteria were adopted to create our survey sections and questions. The survey should include graduates':

1. Family information, demographics, and personal data, such as gender, hobby, language, etc., to recommend personalized recommendations, forming the "Graduate personal information" section.
2. High school or vocational school data, such as graduates' school courses interests, school sector, school education system, etc., to recommend to high school students' recommendations based on their high school information, forming the "Graduate high school or vocational school information" section.
3. University information, such as teaching effectiveness, university major, university name, etc., to recommend to high school students recommendations related to university paths, forming the "Graduate first and currently attended university information" sections.
4. Career information, such as their current occupation, career interests, etc., to recommend to high school students career choices related to their career interests, forming the "Graduate interests and career information" section.

All these criteria have covered the graduates' trajectories, starting from studying at high school, then studying at the university, followed by entering the career market. Integrating university graduates' trajectories data in our hybrid recommendation process would help recommending to high school students promising university paths and career choices. Figure 2 shows some samples of the questions in the conducted survey. The university graduates rated their level based on 23 high school

YOUR INTEREST AND CAREER INFORMATION

What is the highest level of education you have completed? *

Bachelor degree
 Master degree
 Doctoral Degree

What kind of job/career interests you? *

Business, Management, and Administration

Choose the category of your current job? *

Information Technology

What is your current job? *
 Example: Mechanical Engineer, Accountant...
 IT

YOUR HIGH SCHOOL OR VOCATIONAL SCHOOL INFORMATION

What high school did you attend? *
 Multiple answers are allowed

Private school
 Public school

Do you have a high school degree or equivalent certificate? *

Yes
 No

If yes, what is the type of the degree you graduate with from High School?
 Multiple answers are allowed

Technical
 General
 Autre : _____

How would you rate your high school level on the following?
 In this section, please rotate your device to landscape to see all options.

Humanities *

	Very Good	Good	Poor/Not concerned
Literature	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Philosophy	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Religion	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Fine Arts *

	Very Good	Good	Poor/Not concerned
Music	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Theatre	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Dance	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Drawing	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

How important were these criteria when you chose your major/university? *

	Very important	Moderately important	Not so important
Financial criterion (Registration fees, quality of life, housing, ...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geographical criterion (big city, distance home-university, ...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reputation (rank of university, university-company relationships, ...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Family pressure (my parents want that)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Future employment opportunities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 2 Samples of survey questions

courses, namely: Arabic language, Biology, Chemistry, Dance, Drawing, Economics, English language, French language, Other foreign languages, Mathematics, Geography, History, Music, Literature, Physical education, Philosophy, Science of engineering, Physics, Psychology, Technology and Computer Science Religion, Sociology, and Theatre.

4 The investigated approaches

In this section, we investigate five approaches to choose the most accurate one to recommend a major and career domain to high school students based on the trajectories of university graduates. A case study is formulated based on the dataset

collected from the survey conducted in Lebanon. The selection of these approaches was affected by the data types and high dimensionality of attributes in the dataset. These approaches are:

1. The stand-alone user-based and item-based CF RS.
2. The stand-alone DF RS.
3. The stand-alone KB RS supported by CBR.
4. The stand-alone KB RS supported by ontology and CBR.
5. The KB Hybrid RS combined with the user-based CF and supported by ontology and CBR.

The overall research process for this study is represented in Fig. 3. Our dataset has been collected from a survey as presented in Section 3. Then, it has been analyzed and stored into the knowledge base (ontology), whereas the ratings were stored into databases. Each of the five considered approaches implements one or more RS techniques, as shown in Fig. 3. Applying these approaches on our dataset has generated different results of recommendations. We compared these results and evaluated them in order to determine the most appropriate approach for our research questions and to present the optimum personalized recommendations to the high school students. Thus, this study provides a comparative analysis considering two main perspectives:

- The integrity of all data related to the adopted research questions, which is reflected in the four data categories collected in our dataset, as discussed in Section 3.
- Considering all the evaluation criteria that can be applied to evaluate the five recommender systems and support our investigation to answer the adopted research questions, including the Mean Absolute Error (MAE) and Root Mean Squared

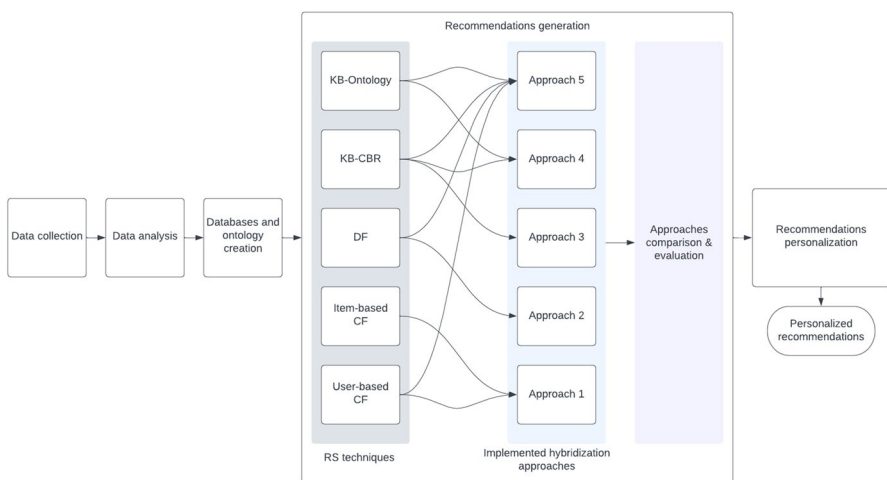


Fig. 3 The overall conducted research process

Error (RMSE) accuracy metrics, in addition to other similarity rate metrics as further explained in the upcoming sub-sections.

A detailed experimental evaluation is demonstrated in the following sub-sections for each approach to examine the accuracy of its recommendations for high school students.

4.1 The stand-alone user-based and item-based CF RS

This approach integrates ratings data to find interests' similarities between high school students and university graduates, so that to generate career domain recommendations. In our case study, the CF recommendation is based on high school students' and graduates' course ratings. The university graduates rated their level on 23 high school courses. Table 1 shows a sample of course ratings for a high school student. The CF RS engine uses these ratings to recommend to him/her a career domain based on the prior university graduates' ratings.

Since the experimental evaluation is based on course ratings, we developed the memory-based technique of CF, in which both the user-based and item-based methods are adopted (Ghazarian & Nematbakhsh, 2015). The user-based method associates similar users to the active user, recognized as neighbor users. Furthermore, missed ratings are predicted using various similarity metrics. The metrics calculate the similarities values based on the past users' ratings. The item-based method focuses on the items instead of the users to find the most similar items based on the active user's ratings compared to the past users' rating history. As for the CF experiments, we used a dataset of 469 objects having 39 attributes. The objects represent the graduates of the university and the 39 attributes represent their high school courses and career ratings. The dataset contains 11,000 ratings for 39 attributes, provided by the 469 graduates. All the university graduates in the dataset rated at least 20 attributes.

Table 1 An example for a high school student courses' rating

<i>Course Name</i>	<i>Evaluation (Very Good=3, Good=2, Poor=1)</i>
Literature	2
Philosophy	3
Religion	2
Music	3
Theatre	1
Dance	2
Drawing	2
Biology	1
Chemistry	3
Physics	3
Mathematics	2

The dataset was split into two sub-datasets in our experimental study; the training and testing data. The Euclidean distance similarity, City Block similarity, Spearman Correlation similarity, Pearson Correlation similarity, Uncentered Cosine similarity metrics (Bagchi, 2015) have been used to find similarities between the graduates and high school students based on their ratings. For each similarity metric, an evaluation has been conducted based on the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) accuracy metrics (Bagchi, 2015). Some parameters, such as the N neighborhood size, and the training ratio, have been defined as shown in Fig. 4. The N neighborhood represents the nearest neighbors to the object location. As for the user neighborhood, the RS can find the most similar users to the selected user. The size of the neighbor can affect the prediction quality. By changing the number of neighbors, the sensitivity of the neighborhood is determined. The training ratio represents the percentage of each user’s preferences to use for recommendations production; the rest of the training ratio is compared to the estimated preference values to evaluate the recommender’s accuracy.

To evaluate the accuracy of this CF RS approach, we implemented the mahout evaluation method (Giacomelli, 2013). For each user, 90% of the preferences provided by the given data model were set as training data to generate the recommendations, where the rest of the data was compared against the estimated preference values to see how much the recommender’s predicted preferences match the user’s actual preferences. The return is a score representing how well the recommender’s estimated preferences match actual values. Lower scores mean a better match and 0 is a perfect match. The result of many experiments shows that the user-based CF algorithm with the Euclidean distance similarity metric, Neighborhood size equal to 50, and training ratio equal to 0.8 generated the lowest MAE and RMSE values that is 0.45 and 0.58 respectively, compared to the item-based CF algorithm and its similarity metrics as shown in Fig. 4. The following is an example of a recommendation based on the user-based CF approach:

“Recommended university major: Information Technology, similarity rate: 3.0”

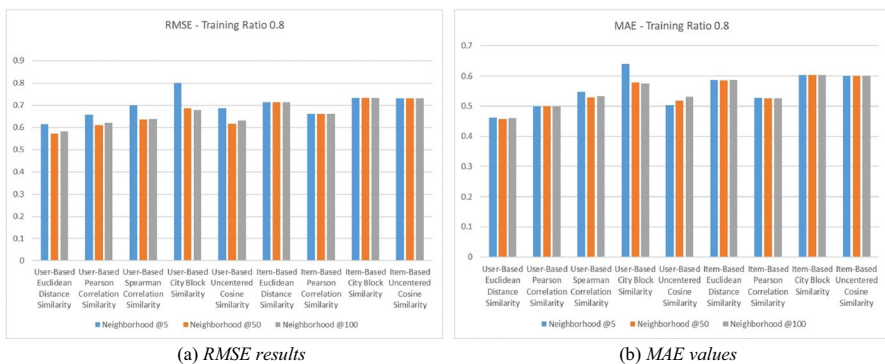


Fig. 4 RMSE/MAE for the item-based and user-based similarities

In the above recommendation, the *similarity rate* represents the highest rate of the recommended major, which means it is 100% similar, whereas *Information Technology* represents the recommended career domain. Even though the obtained results are accurate, this approach was applied to a part of our dataset, which is the rating. However, our dataset contains more heterogeneous data, such as students' interests, career information, and demographic data. Thus, this approach is inadequate for the whole dataset.

4.2 The stand-alone DF RS

The DF RS is based on the demographic data, which does not take into account the domain knowledge, user interests, and ratings in its recommendation process. Thus, the DF RS can provide recommendations before receiving any rating from the active students. However, for many students, generalizations with the demographic features seemed to be too general for the highly personalized recommendations. For example, not all 17-year-old male students who liked scientific courses in high school would prefer the same university major or career in the future. In addition, students with different opinions or unusual interests result in low correlation coefficients with other students. The recommendations for this kind of students are very hard to generate. Thus, the recommendations that are based only on demographic data, such as student's language, gender and location, etc., may lead to inaccurate predictions, leading to the grey-sheep limitation (de Campos et al., 2010).

Therefore, the stand-alone DF RS approach was considered unsuitable for the dataset, since it does not take into consideration the domain knowledge, students' preferences, and rating history. In addition, the experiments revealed that no correlation was found between the demographic data and the courses ratings in the university graduates' dataset. In our case, demographic data are not enough on their own; they must be combined with domain knowledge, course rating, and students' preferences to generate more personalized recommendations. To overcome the problems and limitations of the stand-alone DF RS, different RSs approaches should be examined, such as the KB and hybrid systems.

4.3 The stand-alone KB RS supported by CBR

How to generate personalized recommendations to a high school student based on his/her interests and knowledge and not on his/her courses' ratings? The KB approach could be a good solution to generate recommendations based on the domain knowledge instead of ratings (Tarus et al., 2018). Thus, we implemented the CBR KB approach, which uses indexes to speed up retrievals from a case base. The indexes apply feature matching to organize and label cases so that appropriate cases can be found when needed. Cases may be indexed by an open vocabulary or a prefixed, and within a hierarchical or a flat index structure (Perner, 2019).

As the size of the case base increases, it becomes critical that CBR would access the stored cases efficiently (Recio-García et al., 2014). To address this, jColibri provides a persistence mechanism through different "Connectors" and data structures

Fig. 5 A sample of CBR KB RS query

Preferred Courses	Mathematics
Educational Institution Type	General
Favorite Hobby	Computer
Model as Family Member	true
Career Interests Category	Information_Technology
Ok	

Case ID	Preferred Course	Education	Hobby	Model as a Member of Family	Career Interest	Academic Discipline	University	Occupation
622	Mathematics	General	Computer	TRUE	Information Technology	Computer Science	Lebanese Uni	Information Technology
676	Mathematics	General	Computer	TRUE	Information Technology	Computer Science	Lebanese Uni	Information Technology
56	Mathematics	General	Sports	TRUE	Information Technology	Computer Science	Lebanese Uni	Information Technology
66	Mathematics	General	Sports	TRUE	Information Technology	Computer Science	Lebanese Uni	Information Technology
125	Mathematics	General	Swimming	TRUE	Information Technology	CCE	AUL	Information Technology

Fig. 6 The CBR knowledge-based RS retrieved recommendations

```

))] [Solution: (null;Computer Science;Lebanese university;Information Technology)] [Sol.Just.: null] [Result: null] -> 1.0
))] [Solution: (null;Computer science ;Lebanese university;Information Technology)] [Sol.Just.: null] [Result: null] -> 1.0
Solution: (null;Computer Science;Lebanese university;Information Technology)] [Sol.Just.: null] [Result: null] -> 0.8
Solution: (null;"Computer science;Math";Lebanese university)] [Sol.Just.: null] [Result: null] -> 0.8
]] [Solution: (null;Computer communication engineering ;All ;Information Technology)] [Sol.Just.: null] [Result: null] -> 0.8
    
```

Fig. 7 The CBR KB RS recommendations’ evaluations

for in-memory organization for case base management. jColibri separates the case storage from the indexing structure, where *Connectors* know how to access and retrieve cases from the medium and return those cases to the CBR system in a uniform way. In-memory or indexing is the second layer of Case Base management. *In-memory* case organization is the data structure used to organize the cases once loaded into memory, i.e., linear lists, trees, case retrieval nets, etc. Fig. 5 shows a sample of a high school student’s query via the KB RS’s implemented interface, whereas Fig. 6 shows five recommendations suggested to the high school student based on his/her query. Each recommendation suggests a university/college, university major, and career domain. The high school student can select a suitable recommendation from the retrieved cases that match his/her interests. To evaluate the accuracy of the recommendations, we used *NNScoringMethod* in jColibri2 to measure the similarity rate (Recio-García et al., 2014). This function performs a Nearest Neighbor numeric scoring comparison of attributes to evaluate the retrieval of the most similar cases. Fig. 7 shows that the first two solutions are 100% similar to the above query and the other three solutions are 80% similar to the same query. Although the CBR-based approach provided good results, it does not take into account the semantic similarity between the concepts.

4.4 The stand-alone KB RS supported by ontology and CBR

The KB RS supported by ontology and CBR used career knowledge, higher education knowledge, students’ interests, and demographic data. We constructed “GraduateOnto” as our ontology that encompasses the higher education, school, career, and

student profile concepts as shown in Fig. 8. The purple classes and subclasses represent our GraduateOnto concepts, the green classes represent the DBpedia ontology concepts (Lehmann et al., 2015), and the pink classes represent the Schema.org concepts (Patel-Schneider, 2014). Combining ontology and CBR approaches would improve the personalization of recommendations. We linked our ontology to these sources by reusing some concepts already defined online and thus these concepts are linked to open data. The CBR is based on three elements: *description*, *solution*, and the *case*. Thus, our GraduateOnto and the CBR are connected as follows:

- The graduate concept encompasses the graduates’ *cases* describing all graduates’ instances in the knowledge base.
- The student information, school information, and person concepts cover the *description* of these graduates.
- The university information and the career information concepts encompass the *solution*.

The experiments are based on the prior graduates’ cases stored as instances in our ontology. The graduates’ cases were extracted from our survey based on many criteria, such as using only data related to university graduates having a university major related to their current job and their job that meets their interests. The final refined case base encompasses 658 graduate cases. This case-base is integrated into the ontology design and computed by jColibri2’s *NNScoringMethod* retrieval function

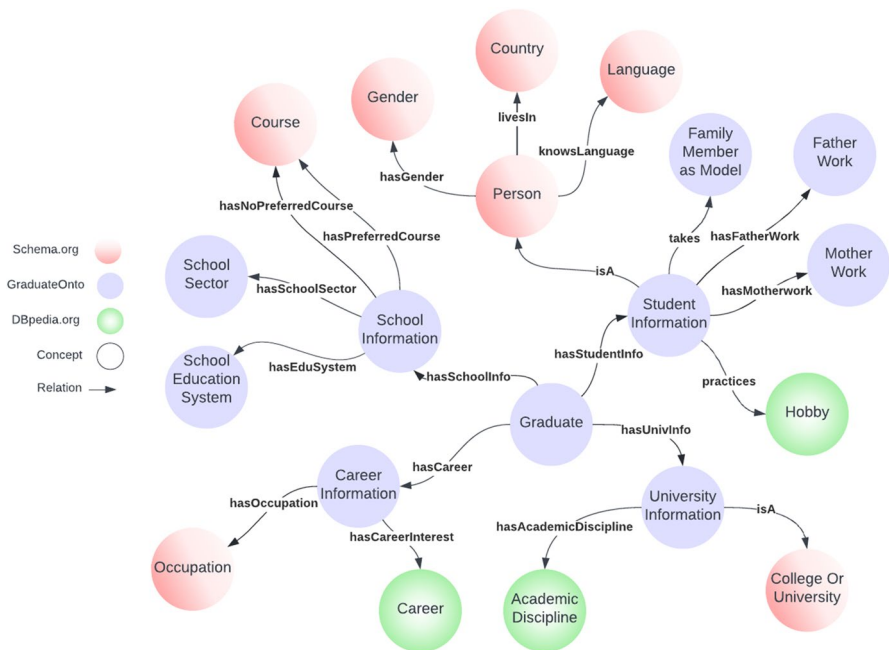


Fig. 8 GraduateOnto design

(Recio-García et al., 2014). It uses: (1) global similarity functions such as the *mean Average* to compare compound attributes and (2) local similarity functions such as *Detail* to compare simple attributes. For example, the Graduate case component is a compound attribute composed of several simple attributes (gender, language, hobby, country, etc.). When two cases are compared, local similarity computes the similarity between simple attributes and global similarity computes the average over the local similarities. Thus, a global similarity function is assigned to the description like the average function. The method returns a collection of *RetrievalResult* objects.

Most similar cases are selected once they have been scored according to their similarity with the query, where only the top *k* most similar cases are selected. Hence, we apply the *k*-NN retrieval process, which combines Nearest Neighbor scoring and top *k* selection. Once the similarity function and weight are set for the attributes, the similarity function is executed to obtain a list of retrieval result objects that contain the most similar cases to the query. Finally, the most similar cases are obtained using the *selectTopKRR* function (Recio-García et al., 2008). We used the *jColibri2* retrieval process to compare high school students' cases with university graduates' cases and find the most similar cases in order to provide appropriate recommendations. Fig. 9 shows an example of a high school student's query. The query's attributes are selected from the instances saved in the ontology design. The scoring of the most similar cases in this process is computed based on the prior cases' similarity with the query. The top *k* most similar cases are retrieved, mixing the Nearest Neighbor scoring and top *k* selection techniques. The calculation returns a value between (zero ~ one), showing the retrieved solution or case being the least or most similar to the active query case respectively. Fig. 10 presents the first retrieved case with 100% similarity to the active user query case in Fig. 9, while Fig. 11 shows the second most similar case to the active user query, which is approximately 88% similar to the submitted query.

This KB RS generated personalized recommendations to the high school student with the support of the ontology and CBR concepts. The *HoldOutEvaluator* algorithm was used to evaluate the accuracy of this RS approach (Recio-García et al., 2014). As shown in Table 2, the evaluation results show high accuracy levels using 10% and 15% of the dataset for testing, carrying out the process several times with a different number of cycles.

Fig. 9 Active student query example

	Value
Preferred Courses	scientificCourses
Your Location	Beirut
Educational Institution Type	vocationalSchool
Favorite Hobby	hunting
Model as Family Member	yes
Career Interests Category	Finance

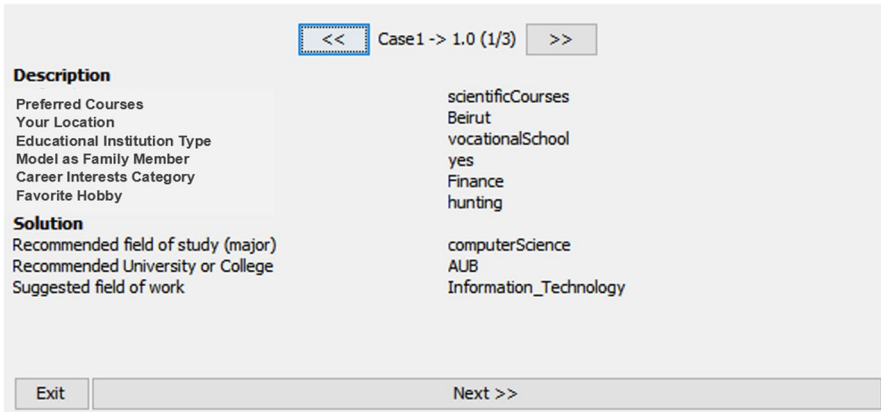


Fig. 10 Most similar retrieved case to an active student’s query

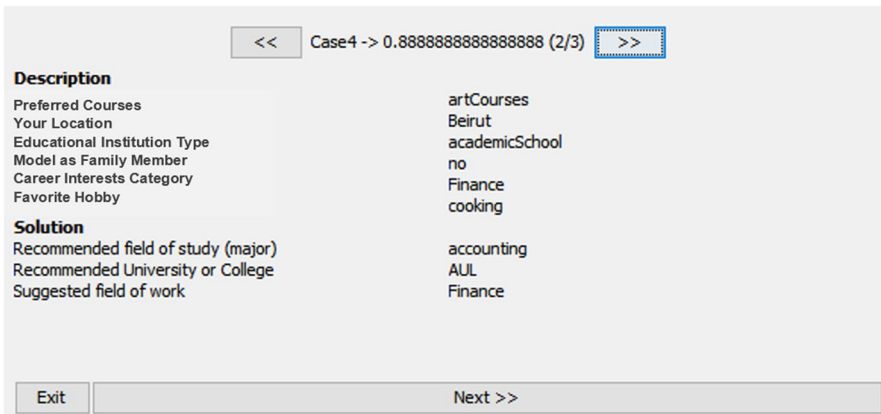


Fig. 11 Second most similar case

Table 2 HoldOutEvaluator evaluation results (approach 4)

	% of dataset for testing	Number of cycles	Time per cycle (ms)	Similarity
Experiment 1	10	65	255.8769	0.96666
Experiment 2	15	98	198.9285	0.96666

Thus, this system is considered more efficient than the previously tested RSs. The analysis of this approach revealed that the ontology is very useful in supporting CBR KB RSs. The ontology helped to integrate the high school students' interests, graduates' knowledge, and high school and higher education knowledge into the KB RS, conceptualizing them in a formal language. Likewise, the reuse of ontology also benefits from its reliability and stability. Moreover, throughout the similarity calculation, the ontology permits linking the gap between the high school student's query and the case-based vocabulary. This integration allowed the system to generate personalized recommendations. However, this approach does not take into consideration the high school students and graduates courses' ratings.

4.5 The KB hybrid RS combined with the user-based CF and supported by ontology and CBR

This approach presents a hybrid RS based on the CF, CBR, DF, and KB techniques with ontology. This approach is a combination of approaches 1, 2, and 4. It allows for generating recommendations based on the domain knowledge, students' ratings, interests, and demographic data. The demographic data of graduates has been integrated into GraduateOnto. Thus, this hybridization aims to improve the recommendations and precision of the system. Fig. 12 illustrates the sequence diagram of this approach. First, the high school student enters his profile, preferences, and ratings of the high school courses into the system's GUI. These data are stored in the corresponding databases. Then, the student uses the search system to get personalized recommendations about the universities, majors, and careers domain. In order to get such recommendations, the user-based CF system interrogates the databases to search for similar users (in our case the graduates) based on the high school courses' ratings. As mentioned in Section 4.1, the recommendation of our user-based CF is based on the

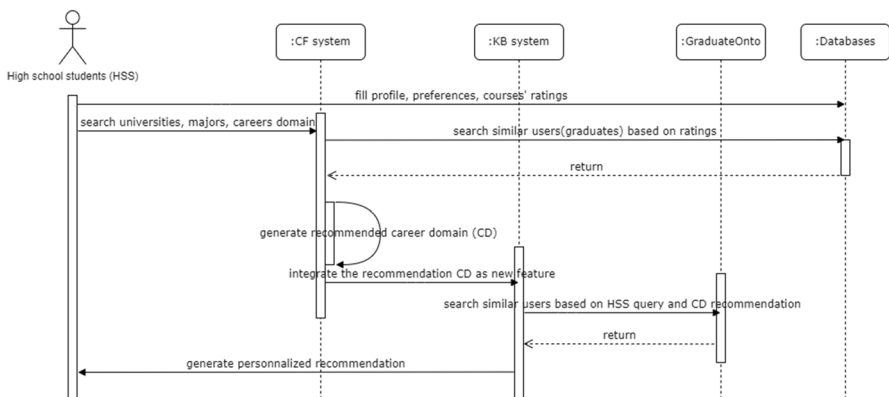


Fig. 12 The sequence diagram of Approach 5

Euclidean distance metric. The CF system selects the most similar result and integrates the career domain information of this result as a new feature in the KB system, by using the Feature Augmentation Hybridization strategy (Bruke, 2002). This new feature is used in the KB system as a assist knowledge to the query of the high school student.

Figure 13 illustrates a query sample requested by a high school student to get recommendations for the university paths, while Fig. 14 shows an example of the “Graduate Interest Career Domain” feature that is integrated with the high school student’s query, as well as the top N recommendations generated as per Approach 5. The KB system interrogates GraduateOnto to search for similar paths between graduates’ cases (that are saved as instances in the ontology) and the high school student, based on his/her search query and the CF recommendation. The system selects the top K results and presents them to the high school student. Approach 5 integrates a dataset that encompasses 658 graduate cases representing only the university graduates that have a university major related to their current job and their job meets their interests. By implementing the ontology similarity and CBR retrieval method, this hybrid KB system can retrieve the most similar cases that best fit the high school student’s interests.

The *HoldOutEvaluator* algorithm was applied to evaluate the accuracy of Approach 5 (Recio-García et al., 2014), and trained with 658 university graduate’s cases as shown in Table 3. Table 4 shows that this approach achieves high accuracy levels as per two criteria, namely the “accuracy of retrieving the most similar cases” and the “accuracy of generating appropriate recommendations”. The tests were conducted with a sample size of 60 high school students, 40 university students, and 40 university graduates. The university students were requested to participate since they have experienced the transition from school to university, whereas the university graduates were requested to participate as they have already passed this transition and know its outcomes. The results are shown in Tables 5 and 6. This test’s purpose is to find out whether the

YOUR PERSONAL INFORMATION	
What is your Gender?	male
Which country do you live in?	lebanon
Select your preferred language	english
Select your favorite hobby	music
Do you take as a model a member of your family?	no
YOUR HIGH SCHOOL INFORMATION	
What high school did you attend? (private or public)	private_school
What is your high school education system? (technical or general)	general
What high school subject did you like best?	mathematics
What high school subject did you like least?	arabic
CAREER INFORMATION	
Select the work domain of your father	science_technology_engineering_and_mathematics
Select the work domain of your mother	business_management_and_administration

Fig. 13 A sample query in Approach 5

use of prior graduates’ knowledge can be applied to assist current high school students. All experiments have proven the efficiency of our user-based CF system supported by CBR, ontology, and KB RS as presented in Table 4. In addition, our analysis indicated that the hybridization in Approach 5 offers the most adequate solution to our high-dimensional data, which include more than 50 heterogeneous attributes.

First similar case result example

<< Graduate584 -> 1.0 (1/5) >>

SIMILAR CASE DESCRIPTION

Graduate Gender	male
Graduate Country	lebanon
Graduate preferred Language	english
Graduate favorite hobby	music
Graduate take as a model a member of his/her family	no
Graduate High School (private or public)	private_school
Graduate school education system	general
Graduate Preferred Course	mathematics
Graduate Not Preferred Course	arabic
Graduate Father Work	science_technology_engineering_and_mathematics
Graduate Mother Work	business_management_and_administration
Graduate Interest Career domain	science_technology_engineering_and_mathematics

RECOMMENDATIONS

Recommended University Field of Study	mathematics
Recommended University or College	aub
Recommended Career Domain	science_technology_engineering_and_mathematics

Second similar case result example

<< Graduate601 -> 0.9583333333333335 (2/5) >>

SIMILAR CASE DESCRIPTION

Graduate Gender	male
Graduate Country	lebanon
Graduate preferred Language	english
Graduate favorite hobby	computer
Graduate take as a model a member of his/her family	no
Graduate High School (private or public)	private_school
Graduate school education system	general
Graduate Preferred Course	mathematics
Graduate Not Preferred Course	arabic
Graduate Father Work	science_technology_engineering_and_mathematics
Graduate Mother Work	marketing_sales_and_service
Graduate Interest Career domain	information_technology

RECOMMENDATIONS

Recommended University Field of Study	computer_science
Recommended University or College	lebanese_university
Recommended Career Domain	information_technology

Fig. 14 Recommendation results in Approach 5

Table 3 HoldOutEvaluator evaluation results (Approach 5)

	% of dataset for testing	Number of cycles	Time per cycle	Similarity
Experiment 1	5	32	217.71875 ms	0.96666666
Experiment 2	10	65	176.98461 ms	0.97757575

Table 4 The accuracy results of Approach 5

Criteria	System accuracy
Retrieving the most similar cases	98%
Generating appropriate top N recommendations based on the students' interests	95%

Table 5 Results of interest in the recommendation of Approach 5

Participants	Total number of participants	Not at all useful	Fairly useful	Very useful
High school students	60	3.333% (2 students)	3.333% (2 students)	93.333% (56 students)
University students	40	2.5% (1 university student)	5% (2 university students)	92.5% (37 university students)
University graduates	40	2.5% (1 university graduate)	2.5% (1 university graduate)	95% (38 university graduate)

Table 6 Results of users' satisfaction in the recommendation of Approach 5

Participants	Total number of participants	Dissatisfied	Satisfied	Very Satisfied
High school students	60	3.333% (2 students)	5% (3 students)	91.7% (55 students)
University students	40	5% (2 university students)	5% (2 university students)	90% (36 university students)
University graduates	40	5% (2 university graduates)	2.5 (1 university graduate)	92.5% (37 university graduates)

5 Discussion and insights

Table 7 summarizes the comparative analysis of the five investigated approaches presented in this paper, showing the advantages and the disadvantages of each approach. The experiments demonstrate that the stand-alone user-based and item-based CF approach is not adequate to our dataset, since it computes only rating history and has many limitations such as cold-start problem, data sparsity and, grey-sheep. Additionally, it recommends only career domains. Similarly, the stand-alone DF approach is not adequate for our dataset since it integrates only demographic data. In addition, our analysis revealed the weakness of DF approach in generating personalized recommendations, as it does not consider the ratings and domain knowledge of graduates. On the other hand, the stand-alone KB supported by CBR is a valuable tool when the item is infrequently used because it is not dependent on the ratings. However, like the first two approaches, this technique is not adequate for our dataset as it does not consider the graduates' ratings. Furthermore, this approach needs to be combined with an ontology-based technique in order to define the graduates' domain knowledge and to retrieve similar cases semantically. This was the aim of the fourth investigated approach, the stand-alone KB supported by ontology and CBR. This approach has generated personalized and accurate recommendations as illustrated in Table 2. However, as the previous approaches, it is not adequate for our dataset as it does not take into account the graduates' ratings. Therefore, it should be incorporated with the CF technique in a hybrid RS in order to compute knowledge and ratings, which is the aim of the fifth approach (KB using DF + user-based CF and supported by ontology and CBR).

Therefore, it can be concluded from this comparative analysis that this hybridization is the most suitable recommendation approach to answer our research questions by generating 98% of similar cases, 95% of them are personalized based on the interests of high school students, since it computes the returns based on the domain knowledge, high school student profile, ratings, interests, and demographic data. The average usefulness of the proposed hybrid approach ranged from 92.5% to 95% as presented in Table 5, whereas the average satisfaction level ranged from 90% to 92.5% as shown in Table 6. Thus, high-accuracy recommendations are generated by integrating the three core DF, CF, KB techniques as presented in Table 4. Furthermore, it shows a high precision in treating heterogeneous data types and high dimensional datasets as shown in Table 3. Based on this discussion and analysis, we recommend using this hybridization approach in other problems, as well as we encourage the reusability of the constructed GraduateOnto ontology in other problems related to the educational domain.

Table 7 Comparative analysis of the five recommender systems approaches

<i>RS technique</i>	<i>Data type</i>	<i>Advantages</i>	<i>Disadvantages</i>	<i>Effectiveness</i>
<i>Standalone user-based and item-based CF</i>	Users' ratings	<ul style="list-style-type: none"> • Capture the updates over time • No need to domain knowledge 	<ul style="list-style-type: none"> • Cold-start problem • Data sparsity • Grey-sheep 	<ul style="list-style-type: none"> • Adequate for a part of our dataset • Recommend only career domains
<i>Standalone DF</i>	Demographic data	<ul style="list-style-type: none"> • Resolve cold-start problem • No need to domain knowledge 	<ul style="list-style-type: none"> • Lack of demographic data • Privacy issues • Grey-sheep problem 	<ul style="list-style-type: none"> • Adequate for a part of our dataset
<i>Standalone KB supported by CBR</i>	Domain Knowledge of prior cases	<ul style="list-style-type: none"> • Valuable tool when item infrequently used • Not dependent on ratings 	<ul style="list-style-type: none"> • Hard to acquire domain knowledge • Do not consider semantic similarity 	<ul style="list-style-type: none"> • Adequate for a part of our dataset
<i>Standalone KB supported by ontology and CBR</i>	Domain Knowledge of prior cases + ontology	<ul style="list-style-type: none"> • Valuable tool when item infrequently used • Not dependent on ratings 	<ul style="list-style-type: none"> • Require experts in domain knowledge 	<ul style="list-style-type: none"> • Adequate for a part of our dataset
<i>KB + user-based CF supported by ontology and CBR</i>	Users' ratings + Domain Knowledge of prior cases + demographic data + ontology	<ul style="list-style-type: none"> • Valuable tool when item infrequently used • Compute semantic similarity between the ontology concepts • Treat high dimensional datasets • Treat heterogeneous datatypes 	<ul style="list-style-type: none"> • Require experts in domain knowledge 	<ul style="list-style-type: none"> • Adequate for the whole dataset (ratings, demographic data, domain knowledge)

6 Conclusion

In this paper, we implemented and evaluated five recommendation approaches in order to select the most appropriate approach to recommend a university, university major (study field) and career domain to high school students based on their profile, preferences, and level at high school. These approaches are: (1) the stand-alone user-based and item-based CF approach, (2) stand-alone DF approach, (3) stand-alone KB approach supported by CBR, (4) stand-alone KB approach supported by ontology and CBR, and (5) KB Hybrid RS combined with the user-based CF technique and supported by the ontology and CBR. In addition, the GraduateOnto ontology is introduced, which describes the domain knowledge of graduates from universities, including their profile, information about their high school studies, university studies, career occupation and interests. The experimental results indicated the efficiency of the KB hybrid RS, combined with the user-based CF and supported by ontology and CBR approach, generating 98% of similar cases, 95% of them are personalized accurate recommendations based on the interests of the high school students. Thus, we deduce that this hybrid approach is promising to guide high school students towards the university paths. The comparative study of the five implemented approaches presented in this paper could help researchers to determine the appropriate hybridization techniques for their work. Furthermore, the introduced combination of CF, DF, KB supported by ontology and CBR is novel and could be a solution for similar problems, regardless of the application domain. Besides, the uniquely constructed ontology could be reused in other problems in the educational domain.

We plan in the future to increase our dataset in order to verify the results on more cases, as well as to extend the ontology by considering more information about the majors/study fields (description, fees, available grants, privileged majors, percentage of unemployment, etc.). In addition, we intend to apply different ontology acquisition approaches, i.e., automatic or semi-automatic construction of ontology, acquiring the corresponding terms and relations between the concepts and embedding them with an easy ontology representation for better retrieval and reuse. Finally, we plan to evaluate the adaptability of our system in another country, such as France, by using a new dataset.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References


- Agarwal, G., Bahuguna, D. H., & Agarwal, D. A. (2017). Solving cold-start problem in recommender system using user demographic attributes. *International Journal on Emerging Technologies (Special Issue NCETST-2017)*, 8(1), 55–61. 7.
- Aggarwal, C. C. (2016). *Recommender Systems : The Textbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-29659-3>
- AlBanna, B., Sakr, M., Moussa, S., & Moawad, I. (2016). Interest aware location-based recommender system using geo-tagged social media. *ISPRS International Journal of Geo-Information*, 5(12), 245. <https://doi.org/10.3390/ijgi5120245>
- Amane, M., Aissaoui, K., & Berrada, M. (2022). ERSDO : E-learning recommender system based on dynamic ontology. *Education and Information Technologies*, 1-13. <https://doi.org/10.1007/s10639-022-10914-y>
- Assami, S., Daoudi, N., & Ajhoun, R. (2019). Ontology-Based Modeling for a Personalized MOOC Recommender System. In Á. Rocha & M. Serrhini (Éds.), *Information Systems and Technologies to Support Learning* (p. 21-28). Springer International Publishing. https://doi.org/10.1007/978-3-030-03577-8_3
- Bagchi, S. (2015). Performance and quality assessment of similarity measures in collaborative filtering using mahout. *Procedia Computer Science*, 50, 229–234. <https://doi.org/10.1016/j.procs.2015.04.055>
- Benouaret, I. (2017). *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*, (Doctoral dissertation, Université de Technologie de Compiègne).
- Bouihi, B., & Bahaj, M. (2019). Ontology and Rule-Based Recommender System for E-learning Applications. *International Journal of Emerging Technologies in Learning*, 14(15).
- Bousbahi, F., & Chorfi, H. (2015). MOOC-Rec: A case based recommender system for MOOCs. *Procedia - Social and Behavioral Sciences*, 195, 1813–1822. <https://doi.org/10.1016/j.sbspro.2015.06.395>
- Breese, J. S., Heckerman, D., & Kadie, C. (2013). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *arXiv:1301.7363 [cs]*. <http://arxiv.org/abs/1301.7363>. Accessed 1 Aug 2022
- Bruke, R. (2002). Hybrid recommender systems. Survey and experiments. *User Modeling and User-Adapted Interaction*, 4, 331–370.
- Budjanovcanin, A., & Woodrow, C. (2022). Regretting your occupation constructively: A qualitative study of career choice and occupational regret. *Journal of Vocational Behavior*, 136, 103743. <https://doi.org/10.1016/j.jvb.2022.103743>
- Capuano, N., Gaeta, M., Ritrovato, P., & Salerno, S. (2014). Elicitation of latent learning needs through learning goals recommendation. *Computers in Human Behavior*, 30, 663–673. <https://doi.org/10.1016/j.chb.2013.07.036>
- Chamandy, M., & Gaudreau, P. (2019). Career doubt in a dual-domain model of coping and progress for academic and career goals. *Journal of Vocational Behavior*, 110, 155–167. <https://doi.org/10.1016/j.jvb.2018.11.008>
- Chavarriaga, O., Florian-Gaviria, B., & Solarte, O. (2014). A Recommender System for Students Based on Social Knowledge and Assessment Data of Competences. In C. Rensing, S. de Freitas, T. Ley, & P. J. Muñoz-Merino (Éds.), *Open Learning and Teaching in Educational Communities* (p. 56-69). Springer International Publishing. https://doi.org/10.1007/978-3-319-11200-8_5
- De Campos, L. M., Fernández-Luna, J. M., Huete, J. F., & Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, 51(7), 785–799. <https://doi.org/10.1016/j.ijar.2010.04.001>
- Deschênes, M. (2020). Recommender systems to support learners' Agency in a Learning Context: A systematic review. *International Journal of Educational Technology in Higher Education*, 17(1), 50. <https://doi.org/10.1186/s41239-020-00219-w>
- Dorotic, M. ., Verhoef, P. C. ., Fok, D. ., & Bijmolt, T. . H. . A. . (2014). Reward redemption effects in a loyalty program when customers choose how much and when to redeem. *International Journal of Research in Marketing*, 31(4), 339–355.
- Duque Méndez N. D., Rodríguez Marín P. A., & Ovalle Carranza D. A. (2018). *Intelligent Personal Assistant for Educational Material Recommendation Based on CBR*. . . Intelligent Systems Reference Library, vol 132. Springer, Cham.
- Farzan, R., & Brusilovsky, P. (2006). Social navigation support in a course recommendation system. *Proceedings of the 4th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, 91-100. https://doi.org/10.1007/11768012_11

- Ghazarian, S., & Nematbakhsh, M. A. (2015). Enhancing memory-based collaborative filtering for group recommender systems. *Expert Systems with Applications*, 42(7), 3801–3812. <https://doi.org/10.1016/j.eswa.2014.11.042>
- Giacomelli, P. (2013). *Apache Mahout Cookbook*. Packt Publishing.
- Gil, A., Rodríguez, S., De la Prieta, F., De Paz, J. F., & Martín, B. (2012). CBR Proposal for Personalizing Educational Content. In P. Vittorini, R. Gennari, I. Marenzi, F. de la Prieta, & J. M. C. Rodríguez (Éds.), *International Workshop on Evidence-Based Technology Enhanced Learning* (p. 115–123). Springer. https://doi.org/10.1007/978-3-642-28801-2_14
- Gomez-Albarran, M., & Jimenez-Diaz, G. (2009). Recommendation and studentsâ? ? Authoring in repositories of learning objects : A case-based reasoning approach. *International Journal of Emerging Technologies in Learning (IJET)*, 4(2009). <https://doi.org/10.3991/ijet.v4s1.797>
- Guarino, N., Oberle, D., & Staab, S. (2009). What Is an Ontology? In S. Staab & R. Studer (Éds.), *Handbook on Ontologies* (p. 1–17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3_0
- Gulzar, Z., Leema, A. A., & Deepak, G. (2018). PCRS: Personalized course recommender system based on hybrid approach. *Procedia Computer Science*, 125, 518–524. <https://doi.org/10.1016/j.procs.2017.12.067>
- Ibrahim, M. E., Yang, Y., Ndzi, D. L., Yang, G., & Al-Maliki, M. (2019). Ontology-based personalized course recommendation framework. *IEEE Access*, 7, 5180–5199. <https://doi.org/10.1109/ACCESS.2018.2889635>
- Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems : Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>
- Jeevamol, J., & Renumol, V. G. (2021). An ontology-based hybrid e-learning content recommender system for alleviating the cold-start problem. *Education and Information Technologies*, 26(4), 4993–5022. <https://doi.org/10.1007/s10639-021-10508-0>
- Jumaa, Y. M., Moussa, S. M., & Khalifa, M. E. (2017). The main aspects of adaptive educational games for normal and disabled/disordered learners: A comprehensive study. In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 348–355, IEEE. <https://doi.org/10.1109/INTELICIS.2017.8260061>
- Klašnja-Miličević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, 56(3), 885–899. <https://doi.org/10.1016/j.compedu.2010.11.001>
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., & Bizer, C. (2015). DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2), 167–195. <https://doi.org/10.3233/SW-140134>
- Maher, Y., Moussa, S. M., & Khalifa, M. E. (2020). Learners on focus: Visualizing analytics through an integrated model for learning analytics in adaptive gamified e-learning. *IEEE Access*, 8, 197597–197616. <https://doi.org/10.1109/ACCESS.2020.3034284>
- Moussa, S., Maher, Y., & Khalifa, M. E. (2020). Learning preferences adaptation based on the Personalized Adaptive Gamified E-Learning (PAGE) model. *International Journal of Intelligent Computing and Information Sciences*, 20(2), 32–52. <https://doi.org/10.21608/IJICIS.2020.48148.1037>
- Patel-Schneider, P. F. (2014). Analyzing Schema.org. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, & C. Goble (Éds.), *The Semantic Web – ISWC 2014* (p. 261–276). Springer International Publishing. https://doi.org/10.1007/978-3-319-11964-9_17
- Perner, P. (2019). *Case-Based Reasoning – Methods, Techniques, and Applications* (p. 16–30). https://doi.org/10.1007/978-3-030-33904-3_2
- Qomariyah, N. N., & Fajar, A. N. (2019). Recommender System for e-Learning based on Personal Learning Style. *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 563–567. <https://doi.org/10.1109/ISRITI48646.2019.9034568>
- Rahayu, N. W., Ferdiana, R., & Kusumawardani, S. S. (2022). A systematic review of ontology use in E-Learning recommender system. *Computers and Education: Artificial Intelligence*, 3, 100047. <https://doi.org/10.1016/j.caeai.2022.100047>
- Recio-García, J. A., González-Calero, P. A., & Díaz-Agudo, B. (2014). Jcolibri2: A framework for building case-based reasoning systems. *Science of Computer Programming*, 79, 126–145. <https://doi.org/10.1016/j.scico.2012.04.002>
- Rodríguez, P. A., Ovalle, D. A., & Duque, N. D. (2015). A Student-Centered Hybrid Recommender System to Provide Relevant Learning Objects from Repositories. In P. Zaphiris & A. Ioannou (Éds.), *Learning and Collaboration Technologies* (p. 291–300). Springer International Publishing. https://doi.org/10.1007/978-3-319-20609-7_28

- Romero, L., Saucedo, C., Caliusco, Ma. L., & Gutiérrez, M. (2019). Supporting self-regulated learning and personalization using ePortfolios: A semantic approach based on learning paths. *International Journal of Educational Technology in Higher Education*, 16(1), 16. <https://doi.org/10.1186/s41239-019-0146-1>
- Salam, A., & Fathurrahmad, F. (2021). Student final project recommendation system model using Case-Based reasoning (CBR) method. *Jurnal Mantik*, 5(3), 1535–1542.
- Sandvig, J., & Burke, R. (2005). AACORN: A CBR Recommender for Academic Advising | *BibSonomy*. Tech. Rep. TR05–015. <https://www.bibsonomy.org/bibtex/ab6c3cb8fdf42fea21eef7d1ab8fd748>
- Sarwar, S., Qayyum, Z. U., García-Castro, R., Safyan, M., & Munir, R. F. (2019). Ontology based E-learning framework: A personalized, adaptive and context aware model. *Multimedia Tools and Applications*, 78(24), 34745–34771. <https://doi.org/10.1007/s11042-019-08125-8>
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative Filtering Recommender Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Éds.), *The Adaptive Web : Methods and Strategies of Web Personalization* (p. 291-324). Springer. https://doi.org/10.1007/978-3-540-72079-9_9
- Shishehchi, S., Banihashem, S. Y., Zin, N. A. M., & Noah, S. A. M. (2012). *Ontological Approach in Knowledge Based Recommender System to Develop the Quality of E-learning System*. 9.
- Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, 50(1), 21–48. <https://doi.org/10.1007/s10462-017-9539-5>
- Tarus, J. K., Niu, Z., & Yousif, A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems*, 72, 37–48. <https://doi.org/10.1016/j.future.2017.02.049>
- Xia, W., He, L., Gu, J., & He, K. (2009). Effective Collaborative Filtering Approaches Based on Missing Data Imputation. *2009 Fifth International Joint Conference on INC, IMS and IDC*, 534-537. <https://doi.org/10.1109/NCM.2009.128>
- Zhang, X. (2021). Preparing first-year college students' academic transition: What is the value of complementary web-based learning? *Computers & Education*, 172, 104265. <https://doi.org/10.1016/j.compedu.2021.104265>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Christine Lahoud¹ · Sherin Moussa^{1,2}  · Charbel Obeid³ · Hicham El Khoury⁴ · Pierre-Antoine Champin³

Christine Lahoud
christine.lahoud@ufe.edu.eg

Charbel Obeid
cobeid@liris.cnrs.fr

Hicham El Khoury
hkhoury@ul.edu.lb

Pierre-Antoine Champin
pierre-antoine.champin@univ-lyon1.fr

¹ Laboratoire Interdisciplinaire de l'Université Française d'Égypte (UFEID Lab), Université Française d'Égypte, Cairo 11837, Egypt

² Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

³ LIRIS Laboratory, Claude Bernard University Lyon 1, Villeurbanne, France

⁴ LaRRIS Laboratory, Lebanese University, Beirut, Lebanon