LRE Sub-Committee White Paper #2

Final draft - June 2019

A review of the state-of-art of the use of Machine-Learning and Artificial Intelligence by educational portals and OER repositories

Dr. Kamakshi Rajagopal
kamakshi.rajagopal@gmail.com

This white paper was commissioned and funded by the Learning Resource Exchange Subcommittee (LRE - http://lre.eun.org) of European Schoolnet (EUN). It is the second in a series of whitepapers exploring pedagogical, legal, and technical aspects of digital technologies in K-12 European Education.

Introduction

Learning resource platforms are platforms where learners and teachers can access, share and use valued learning resources. With the availability of digital communication means, teachers can easily access resources used and valued by their peers, and equally share the resources that they value with others.

The explosion of learning materials has also been propelled by the Open Education movement, and their use and re-use of Open Educational Resources and Open Educational Practices. In a diverse area such as Europe, educational resources platforms have emerged in every member state with different focuses. As the use of the learning resource platforms increases, they can become messy and unstructured spaces where users find it difficult to find the information and/or resources they need.

Learning resource platforms present a unique context:
First, learning resource platforms can deal with different types of data: learning resources can be fully textual, or incorporate video, audio and multimedia. They can refer to educational games or educational practice and design. Furthermore, they can consist of a single item (e.g. an illustrative video) or integrate several items (e.g. a learning module developed around a series of videos with a learning design around it). Consequently, when learning resource platforms look to introduce recommendation, they need to consider the complexity of the platform and the items that are being recommended. Second, learning resource platforms inherently bring together teachers around learning resources. The platforms can therefore potentially have a social network structure, that adds additional complexity but equally potential for recommendations. Third, learning resources platforms that incorporate social networks are not only a storage repository, but platforms for sharing and reuse of learning resources. The extent to which this sharing and reuse is supported by the platform functionalities can also determine the extent to which interaction can be incorporated in recommendation strategies. Finally, with the increased use of mobile devices, the context of use of learning resource platforms also becomes an important factor to consider. Recommender systems, algorithms that present the user with relevant results based on a chosen user profile, could be a welcome addition to learning resource platforms. This technology is designed to deal with an abundance of messy data. This report presents the current state-of-the-art in the evaluation of recommender systems with a perspective to identify opportunities and challenges in using these tools in learning resource platforms aimed at K12 teachers.

Purpose of this report

This white paper aims to guide learning resource platform owners in their evaluation of the use of recommender systems on their platforms, from a users' point of view.

This report:
- Briefly introduces the technology of recommendation
- Identifies the key issues and decision points in determining if a recommender systems is the right technology
- Gives an overview of the factors playing a role in the implementation of recommender systems
- Lists challenges and recommendations in the implementation of recommender systems

Methodology used

The background literature search for this report was gathered in the following way:

A search for the evaluation of the use of recommender systems in K12 brought up few relevant results. A broader search on evaluation of the use of recommender systems in general, resulted in 398 peer-reviewed articles between 2015-2019. Only 24 of there were dealt with issues on user-centric evaluation.

These 24 articles were analysed with four emerging topics:
- The relationship between recommender algorithms and the perceived user experience
- design of the visualization of recommendation results to the user
- frameworks and methodologies for conducting user-centric evaluations of recommender systems
- criticisms of user-centric evaluation methodologies

The report was drafted based on this background literature and a workshop with the author and the EDRENE members.

Context: Recommendation of learning resources

What is a recommender system?

Jean is looking to buy a new lawnmower. She has a sizeable garden, but nothing too big. She is keen on gardening but thinks of mowing the lawn as a chore. She has been looking for the newest lawnmowers and has also found a whole category of robotic lawnmowers. How can she decide which one to buy?
She might ask salespersons at the store to give her advice, but could also ask her friend Mieke hoe is much more at home in gardening.

"Recommender systems assist and augment this natural social process. In a typical recommender system people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients. In some cases, the primary transformation is in the aggregation; in others the system's value lies in its ability to make good matches between the recommenders and those seeking recommendations."
(Resnick & Varian, 1997).

A recommender system is a piece of software that can look into the relationship between users and items to recommend particular items to a particular user. The system, in some way, uses user information to rate, select and present relevant products, items or information to the user. Several techniques can be used to construct these ratings and rank the selection.

Recommender systems are available for all types of items resources, ranging from physical objects (shoes, furniture and houses) over experiences (travel, social activities and gyms) to virtual objects (music, movies, news, text).

Jean is looking on commercial platform for electric robotic lawnmowers. The system presents a list of lawnmowers that could be interesting for her. The system indicates why this particular item on the list may be useful for Jean.
Several of the recommended items have been bought by other users who were also looking for lawnmowers. There are also lawnmowers that are electric but not robotic. There is even an robotic vacuum cleaner in the list. At the top of the list is a robotic lawnmower from the brand "MowGrass" for a surface area of 10-30m2. It is also mentioned that her friend Mieke has bought a lawnmower of the brand "MowGrass". Jean decides to buy this lawnmower.

General architecture of a recommender system

Every recommender system has certain high-level components: user profiles, item databases, an recommendation engine and a ranking mechanism. Figure 1 illustrates these components and the relationship between them.

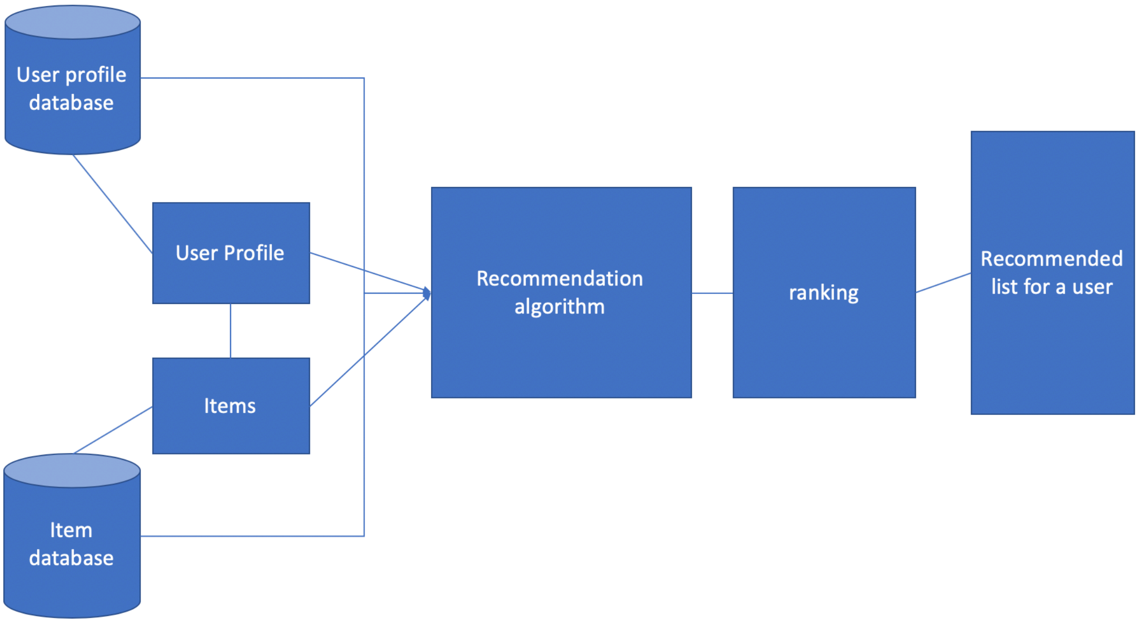

Figure 1 General architecture of a recommendation engine

A recommender system starts from two databases: a database of user profiles and a database of items. There are links between both databases, expressing the relationship between a user and an item. This relationship can be one of interest expressed (e.g. a "like") or one of engagement (e.g. a purchase).

A recommendation is created for a particular user in the database. Starting from one user profile and the related items in the database, a recommendation algorithm calculates which other items in the database might be of interest to the selected user. These algorithms can use different strategies to calculate a recommended list. The most commonly used algorithms are content-based recommendation and collaborative filtering.[1]

a. Content-based: Content-based recommendation technique is closely linked with supervised machine learning that recommends an item to a user based on a description of the item and a profile.
b. Collaborative filtering: Collaborative Filtering techniques use a database for preferences of items by users to predict additional topics or products a new user might like. This type of recommender strategy is divided into (i) memory-based; (ii) model-based and (iii) hybrid strategies.

[1] See this for more examples of https://www.slideshare.net/tommasodinoia/tutorial-recommender-systems-meet-linked-open-data-icwe-2016-lugano-07-june-2016-v11

c. Hybrid Recommender System: This combines different strategies and techniques to overcome limitations of separate techniques

Additionally, some techniques can be used on all recommendation strategies to create the recommended item list and influence the final ranking of items. These techniques are not mutually exclusive and can occur together.

1. Context-awareness: the design can take into account the real-time context of the user, and the user's behavior in this context. This can include information on time, location, setting (work, home, outside), etc.
2. Social graph-based recommendation: information from the social graph in which the users are located, is used to determine the rating and selection of items
3. Diversification: The design can promote more diverse or less diverse items in the final recommended list.

A general issue that comes up at the start of a recommendation engine, in particular with collaborative filtering strategies, is the "cold start" problem. This refers to the sparseness of data relating user profiles to items due to the limited use of the system. With increasing use, the recommendation system can become better.

Search is not recommendation, which is not adaptation

Recommendation is just one technique that can be used to introduce new items of value to a user on a learning resource platform. In this section, we briefly look at the difference between recommendation and other functionalities, in particular, search and adaptation.

Search
Search is most useful when the user is looking for something specific and needs an answer to a specific question. User has free choice in looking for different things. A search functionality will return the best possible answer to the question asked. To quicken response and increase accuracy, search functionalities can give users the ability to limit the search area, for example, why indicating key areas of interest, location, category, etc.

Adaptation
In adaptive systems, the system presents itself in a particular way depending on the user. This means user X might see and experience a different system than user Y based on their previous activity on the system. From a user perspective, they might not know of the difference, as they have their personalized interface and system-world. The user has no free choice in how the system presents itself, and cannot remove themselves from the system to understand how it works.

Recommendation
In recommendation systems, the user has expressed interest in a particular topic or product through a search, purchase or some other form of engagement. Recommendation extends that interest by suggesting related topics or products to the user. The user has a free choice in selecting particular recommended items, which then adds to the user history on the platform and their user profiles. Recommendation is usually just one functionality in a larger platform and therefore does not (necessarily) create a whole system-world experience. Determining what "related" means in this context is the focus of recommendation strategies.

Choosing to implement a recommender system: Things to consider

Do learning resource platforms need features of recommendation? In this section, we look at factors that might determine the choice to implement such a system in the platform.

User behavior on the platform

Recommendation offers solutions for two user needs: exploration and improved search. Starting with improved search, recommendation strategies can be used in cases where users are looking for something specific. Considering information about social connectedness on the platform may provide better ranking of the recommended search results, thereby improving the search engine strategy.

But recommendation comes into its own when it can extend the user's search by introducing the user to information that they might not have intentionally looked for, but is indeed of interest and value to them. This creates a feeling of satisfaction on the part of the user on their activities on the platform.

As a learning resource provider, it is therefore very important to understand the motivations of the users of the platform and especially on the following questions:

- Why do your users come to the platform?
- Which content questions do they have?
- How do your users use the platform? Do they search to find or explore to discover?

Much of the user activities depend on at which stage in their educational material development they reach out to a learning resource platform. Teachers and educators might engage with the platform differently, depending on how they design their classes and courses.

Learning platforms might also allow the sharing of learning resources. In this case, teachers do not only use the platform to find resources but are also creators who contribute and store material that they find valuable.

> LRPlatform is a learning resource platform for teachers and educators in secondary school.
> John is an English teacher for K11-12 and often uses LRPlatform to find good video material to use in his classes. He has a well-designed course for these years, and uses the platform to find the right videos to complement his existing course. The search function is the most important feature for him.
> Francine teaches Mathematics in K8-9 and often looks for new ways to engage her pupils. The recommendation system on the LR Platform often presents interesting approaches in maths, but also physics and chemistry to introduce complex mathematical constructs to her students.

Understanding the user's activities on your platform will enable you to decide the appropriate use of recommendation.

Building a complete user experience

Learning resource platforms are complex platforms that offer a multitude of services for their users:

1. They are repositories of good quality resources – with a possible quality assessment
2. They are spaces where users can search for and find valuable resources
3. They are spaces where users can have continued, dependable access to good resources
4. Some are spaces where users can store their own resources and share them with peers.

Within this context, it is important to remember that recommendation can fulfill multiple roles. Consequently, any recommendation strategy that is used needs to be tied in with the complete user experience that the learning platform is trying to offer. Recommendation is one part of the services a learning platform can offer.

Some questions that need to be considered regarding this are:

- What user demand can recommendation fill?
- Which types of recommendation might be valued by the user? (e.g. of items; of peers; of new items; etc.)
- What is the complete user experience that my learning resource platform offers to a teacher or educator?
- How does any recommendation strategy fit in with the other services that platform provides?

Thinking in terms of the complete user experience will allow the platform provider to specify the role of recommendation within the scala of functionalities offered.

Francine teaches Mathematics in K8-9 and regularly uses LRPlatform to discover new resources to use in her class. She has discovered some very useful resources that way, and regularly logs in on the platform to look at the new recommended resources.
She has also seen the recommended social contacts on the platform, of other Maths teachers and educators using the same resources. However, she has not used this list to connect with others on the platform. For her, LRPlatform is a space of inspiration and not a social platform.

Measuring success

Platform providers might have different goals with an implementation of a recommender system on their learning resource platform. To see if these goals are reached, it is important to have a good view on what these goals are and how they can be measured.
In broad lines, we distinguish three areas of measures: user presence, user engagement and quality assessment. The measures can be situated on a platform level, a recommender system level and a user level (Table 1).

Table 1 Measures for determining success

	User presence	User engagement	Quality assessment
Platform	Duration of startup Number of users Number of returning users	Average time spent on platform per session Number of social activities on platform Social network measures (density, centrality, etc.)	
Recommender system		Click-through rate (CTR) Conversion rate	
User	Periodic survey	"likes" Favorited resources	"likes" Favorited resources

Some measures are general across the platform, and primarily relate to time spent (effectively) on the platform and numbers of users reached. An important measure to keep in mind is duration: recommender systems need some time to become an effective input on a platform (related to the cold start problem described above). Moreover, the more a recommender system is used, the more effective it can become. From a platform provider's point of view, it is important to decide when the effectivity of the recommender system needs to be on point, and how much time the provider is willing to give for this. Additionally, it may pay to have accompanying strategies to grow engagement on the platform during the start-up of the recommendation engine.

Important measure related to engagement are:

- click-through rate (CTR): measure defining the rate at which users click on recommended items to gain more information on the item
- Conversion rate: measure defining the rate at which users not only click on the recommended item, but also engage with it in some way (e.g. purchase, download, effective use, etc.)

If the platform is built as a social network, various social network measures might be useful to measure engagement over time. Finally, explicit user feedback can also be an important source of information on how effective a recommender system is.
Measuring success can be extended to the whole user experience. In particular, the diversity of functionalities provided and how the functionalities might work together can be evaluated

through a user experience perspective. Platform providers should spend some time in considering and defining the measures of success that matter to them.

Implementing a recommendation system: Things to consider

In this section, we will look at the practical implantation of a recommendation system on a learning resource platform.

Starting up: Roadmap

If you have decided that a recommender system will fulfill a need for the users of your learning resource platform, it is useful to gather all this information in a plan. This section provides an overview of the steps you need to take to design your recommender system.
The roadmap presents things you can do to understand your users' needs and the current user experience.
The template of user experience design can help you in defining the features of the new user experience you want to build.

Roadmap

1. Investigate the current user behavior on the platform
    a. Gather some basic statistics on user presence and user engagement
    b. Organize a user focus group interview to understand what users need from your learning platform
    c. Conduct an initial user survey to understand how your users engage with your platforms and what they value in your platform.
2. Identify to what extent the current user experience fulfills the user needs.
    a. What does the current design do well?
    b. Where are the gaps between user needs and user experience?
3. Define the new user experience
    a. What is the interaction between the different functionalities?
    b. Which role does recommendation play here?
4. Define your success measures for the recommendation strategies
5. Define how you will continuously monitor the reception of the recommendation engine as part of your business processes

Table 2 presents a template you can use to describe the current and desired user experience, including the interaction between functionalities. You can also describe the business processes used to track the response to the implemented innovations.

Table 2 Template for establishing current user behaviour and desired user experiences

User behaviour		
Designed Complete User Experience		
Functionality	Purpose	Measures of success
Search		
Recommendation of people (peers; experts; others)		
Recommendation of resources		
Evaluative business process		

Lessons learnt on Design of Recommendation

Much research has been conducted on recommender systems in various fields. This research has primarily focused on the design and performance of algorithms. Some research has also looked into user-centric evaluation of recommender systems. In this section, we will look at the lessons learnt on user appreciation on three aspects: 1) regarding the recommender system and its components; 2) regarding user interaction and user experience; 3) regarding user-centric evaluation.

Recommender system and components

User-centric evaluation of recommendation has brought up several lessons regarding the algorithms used and the user interface. Below, we present some relevant research and related findings.

Algorithms

Research on recommendation algorithms often take a comparative design where two algorithms are ranking filters are set up against each other. These studies start from a technical evaluation of recommender systems, focusing on the extent to which the algorithm ranks the best matches in a dataset for a particular user. For this purpose, the measure of accuracy of the recommended items is the primary measure of effectiveness of the algorithm. The chosen studies however go beyond this technical measure to include other measures indicating user satisfaction. For a detailed description of the measures, we refer to the individual studies.

<u>For content-based techniques</u>

De Pessemier et al (2016) look at context-awareness in a content-based recommender system of news items (De Pessemier, Courtois, Vanhecke, Van Damme, Martens & De Marez, 2016). Taking a content-based strategy had a positive effect on explicit user feedback (in the form of a "thumbs-up") compared to prior-defined explicit static user preferences on news topics. The inclusion of context-awareness (in the form of time awareness) made little difference in this research.

Jannach et al (2015) look at movie recommenders using content-based technique against a non-personalized popularity-based baseline (PopRank) (Jannach,, Lerche & Jugovac, 2015). In this research, content-based recommendations were perceived to be diverse. The perceived surprise factor for the popularity-based methods was low. On the other hand, users indicated that transparency is high when popular items were presented (both in non-personalized personalized way).

<u>For collaborative filtering techniques</u>

Willemsen et al. (2016) look at a movie recommender based on collaborative filtering with an additional focus on diversification (Willemsen, Graus & Knijnenburg, 2016). In this research, participants expect it to be easier to choose from a recommended set with higher diversity and attractiveness. Diversification is also potentially beneficial in reducing effort and difficulty without negative effects on overall satisfaction with chosen item. However, choice satisfaction also depends on the length of the recommended item list: it is low in a short list and high in a long list of recommendations.

Ferwerda et al. (2016; 2017) have looked at music recommenders based on collaborative filtering with a diversification algorithm (Ferwerda, Graus, Vall, Tkalcic, & Schedl, 2017, April; Ferwerda, Graus, Vall, Tkalcic & Schedl, 2016, August). Their findings present a complicated picture.
When users perceive a high level of diversity, it has a positive influence on discovery (enriching user's taste). List familiarity negatively influences discovery. Perceived diversity in the recommended list has negative effect on the attractiveness of the recommended items, and this perception gets stronger for those with a predefined preference. However, perceived diversity has a positive effect on discovery. Users indicate that choosing is easier when

recommendations are deemed more attractive. They also indicate that preference strength and list familiarity are positive for choice simplicity. In short, if users know what they are looking for, diverse lists are not attractive.
The picture becomes more complex when considering personality traits of users: conscientious people judged a higher degree of diversity more attractive and were more satisfied with it. Agreeable people scored list attractiveness and satisfaction higher in a medium degree of diversity. In short, the extent to which diversity is appreciated by the user also depends on the outlook of the user.

Fazeli et al. (2018) investigated recommender systems of learning resources and looked at graph-based memory-based recommendation against collaborative filtering strategies (Fazeli, Drachsler, Bitter-Rijpkema, Brouns, van der Vegt, & Sloep, 2018). The graph-based recommender received a somewhat larger average rating score for perceived usefulness, novelty and serendipity of the recommendations by users.

For hybrid filtering techniques

Ochirbat et al (2018) looked at a hybrid recommender system which integrates content-based techniques into a collaborative filtering system (Ochirbat, Shih, Chootong, Sommool, Gunarathne, Wang, & Ma, 2018). This recommender system supported students in choosing for occupation/job/ study major counselling. This study used several measures to calculate user satisfaction.

This overview of studies shows that user satisfaction can be measured in several different ways, and comparing algorithms is not straightforward. The multitude of additional options such as context-awareness and diversification make that user satisfaction does not necessarily discuss the same issues in different settings. Some differences between algorithms can be noticed but no conclusive decisions can be taken. However, the primary conclusion here is that although algorithms are important, they form only a small part of what defines the final user experience.

Visualisation

An equally important part is how the calculated recommended items are presented to the user. This has a great influence on the user engagement and the resulting user experience (Arana-Llanes, Rendón-Miranda, González-Serna & Alejandres-Sánchez, 2014; Karga & Satratzemi, 2019).

User interfaces can be varied and offer many different options to the designer to interact with the users. Literature shows that much thought needs to go into how recommendations are presented, and how this influences user engagement and experience, and ultimately, user behaviour.

Considering the use of language, recommendations can be presented in neutral language, or pre-structured into categories. These categories, determined by the designer of the recommender system, can influence the perception of the recommendations (e.g. pros vs cons in Karga & Satratzemi, 2019).

A visual representation of the recommended results could give users more control over their recommendations, thereby creating more opportunity for engagement. Parra & Brusilovsky

(2015) present a user controllable personalized system, SetFusion. Its visualisation interface contains three main components: (a) sliders to adjust the importance of each recommendation method, (b) an interactive Venn diagram, and (c) ranked list of recommendations. This is illustrated in Figure 2.

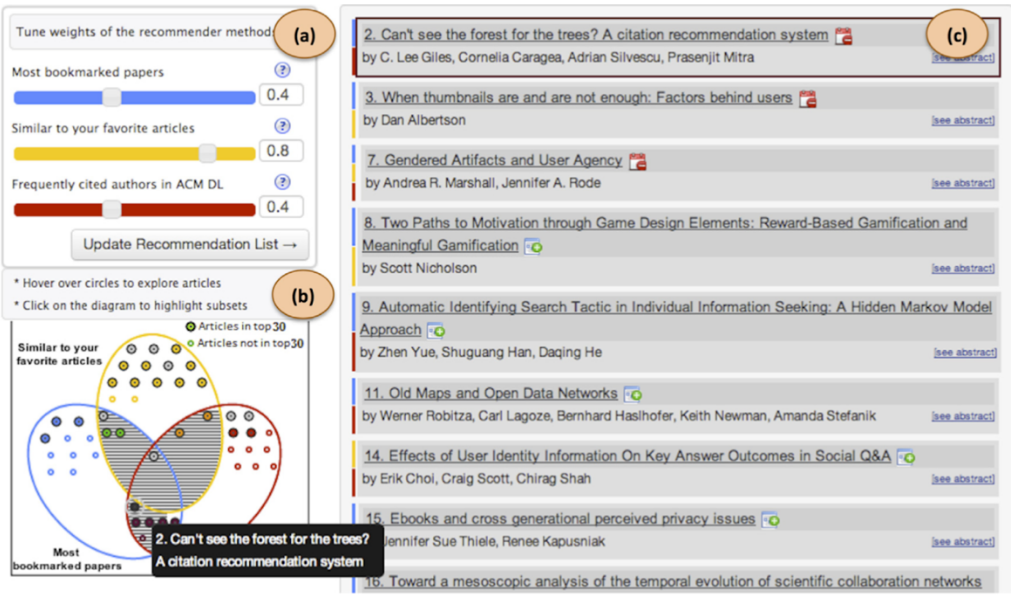

Fig. 1. Screenshot of SetFusion interface. The three components indicated are (a) sliders, (b) Venn diagram, and (c) list of recommendations.

Figure 2 User-controllable user interface of recommendations from Parra & Brusilovsky (2016)

The user-centric evaluation in Parra & Brusilovsky (2016) of these visual interfaces are encouraging:

- users appreciate the controllable recommender interface, but this is mainly when they have experienced a non-controllable one
- a controllable interface seems to give users a better understanding of the recommender system compared to a baseline interface. This seems to be strongest when users have experienced a non-controllable interface.
- the general trusting propensity of a user plays a role in the engagement with the user interface (e.g. high trusting propensity led to the increased perception of the Venn diagram as a useful tool and to an increased perception of the Venn diagram as a tool that increased trust on the recommendations).
- the past expertise of the user plays a role on different levels

Giving the user more control over the steps in the recommendation process, including the option to "backtrack" or reconsider choices made before, seems to be an effective way of engaging the user, as illustrated in the constraint-based recommender system in Epifania & Porrini, 2016.

Other strategies to give user controls over the recommendations they receive, include asking users' explicit and implicit feedback on the effectiveness of recommendation techniques (as in Said, De Luca, Kille, Jain, Micus & Albayrak, 2012). This allows to build a user profile that includes user-preferential recommenders.

For visualisation of recommendations: Etemadpour, Linsen, Paiva, Crick & Forbes (2015) present a task taxonomy to visualise multidimensional data projection. They distinguish four task types: Pattern Identification Tasks; Relation-seeking Tasks; Behavior Comparison Tasks; and Membership Disambiguation Tasks.

User interaction and user experience

In the section above, we have mainly discussed the recommender system and its components. However, the user experience on a learning resource platform is determined by more than the recommendations the user might receive.
As mentioned before, a learning resource platform will have several functionalities that work together to create a user experience. One part of this is the extent to which the user interacts with the different features available.

Taking the visualization of recommendations as a example, we can see that presenting the recommended items as a list does not give users much insight into where the recommended items come from nor the reasoning on why they are recommended to them personally. A more structured visualization including explanations and reasonings might give users more insight into the works of the platform and thereby engage them more.

Even more engaging is giving the user control over the recommended list. Providing options for the user to set preferences in recommendation, to tweak the level of diversity in the recommended list, or to switch on social recommendation (which takes into account the user's personal social network) will give users a perception of more control over what is recommended to them. Building on top of a good database of quality material, this will engage users again to interact more with the platform.

Other ways to create a more unified user experience is to think explicitly how the different components of the learning resource platform can work together to be of value to the user. Search, recommendation, exploration and social network offer different opportunities to engage the user.

LRPlatform is a learning resource platform for teachers and educators in secondary school.
Terese is a proficient user of the LRplatform. She started using the platform about 5 years ago and keeps going back because of the good material that she finds there.
As a Physical Education teacher, she regularly searches for videos on good learning designs for group interaction on the platform. She knows from experience that her search should include certain terms (e.g. “lesson plan” is better than “lesson design”) to get the best results.
About two years ago, LRPlatform introduced a recommendation engine, that offered personalized lists of videos on PE classes. In the beginning, the recommendations were not that good, but this has improved over the years.
Through the recommender system, she has discovered a new concept of interdisciplinary co-teaching where teachers of different disciplines create lessons together, around a common topic, but with diverse foci. She also discovered several colleagues in different schools who are also interested in this concept.
The recommender system on LRPlatform was changed a while ago to include different mechanisms. One of the new features on the platform is that she can decide which algorithm to use and also tinker with the parameters of the algorithm. She spent some time playing around with these settings to see how it changes her personalized recommendation list.
Sometimes when she doesn’t see interesting links in the recommended list when she logs in on the platform, she tweaks the settings a bit and usually finds something of interest.
She always plans to spend 10minutes on the platform, but quite often ends up browsing around for half an hour or even more!

User-centric design and evaluation

User-centric evaluation of recommender systems focuses on the extent to which the user engages with the system, and finds what they are looking for. However, the term "user-centric" can be problematic. It encompasses different meanings in practice.
For example, taking user data into the recommendation algorithm to improve accuracy of the recommender system is deemed user-centric. Using implicit user feedback through likes, or the business measures - click-through rate (CTR) and conversion rate (CR) - to gain insight into what user is attracted to is also considered user-centric evaluation. User-centric evaluation can also be asking for explicit user feedback through the use of surveys, questionnaires, etc. on individual recommended items or recommended item lists.

"Subjective valuations—gathered through psychometric questionnaires—typically provide a more robust measurement of users' experience than behavioral measures, and are better predictors of longer-term system goals such as adoption and user retention."

Knijnenburg, 2016.

Literature shows two primary models for user-centric evaluation, the ResQue model of Pu & Chen (2010) and the user-centric evaluation framework of Knijnenburg, Willemsen, Gantner, Soncu & Newell (2012). Below, we present both these frameworks in general terms, to introduce some concepts of user-centric measures.

Pu & Chen (2010) describe a validated user-centric framework, consisting for 4 layers of higher-level constructs (Fig. 3). Perceived System Quality refers to the users' perception of object characteristics of the system.

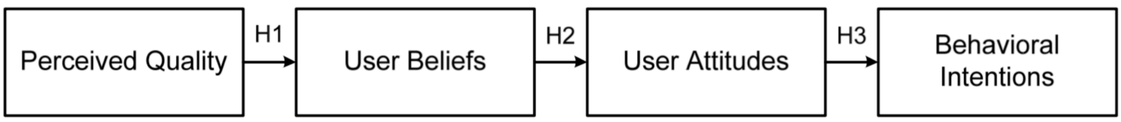

Figure 3 User-centric evaluation framework from Pu & Chen (2010)

Perceived System Quality refers to the users' perception of object characteristics of the system. This includes subconstructs such as Recommendation Quality, Interface Adequacy, Interaction Adequacy, Information sufficiency and Explicability. This construct positively feeds into the users' Beliefs in the recommender system, such as perceived usefulness, perceived ease of use, user control and transparency. This again feeds positively into the User Attitudes towards the recommender system, encompassing overall satisfaction and confidence and trust in the system. Building on these three layers, the fourth layer of Behavioural Intentions say something about the users' intentions to engage with the recommended content (either through use of the content or intention to purchase).

Another model for user-centric evaluation is that of Knijnenburg, Willemsen, Gantner, Soncu & Newell (2012) (Fig 2.). This model differentiates between Objective System Aspects (OSA) such as the recommendations it provides, the interaction functionalities it provides etc. and the Subjective System Aspects (how the user perceives the system) such as perceived quality and interaction usability. Both of these aspects influence the Attitude of the user towards the recommender system, and their Behavior in using the recommender system. However, user attitude and user behavior are influenced greatly by the Situational Characteristics and Personal Characteristics.

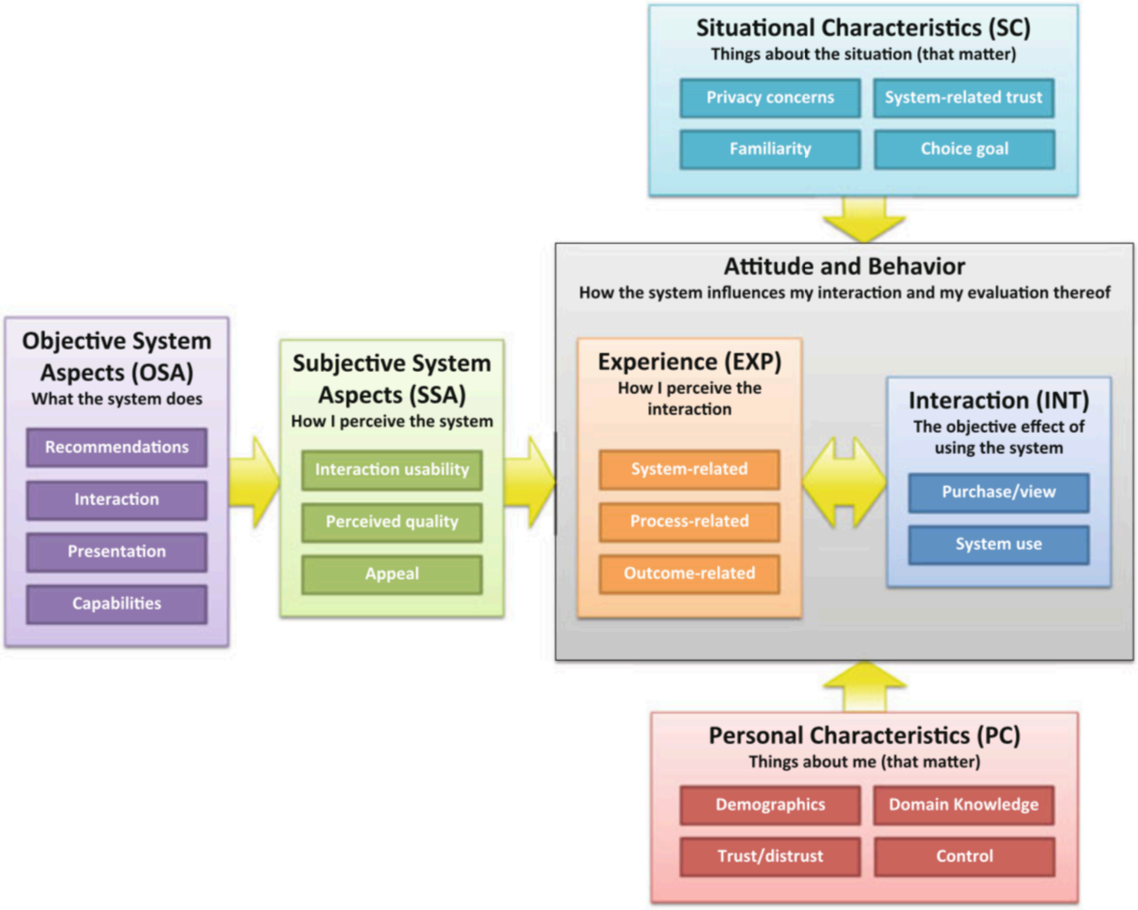

Figure 4 User-centric evaluation framework of Knijnenburg, Willemsen, Gantner, Soncu & Newell (2012)

Shi (2014) introduces some practical instruments and scales for the evaluation of social adaptive e-learning system, to achieve high quality Learner Experiences. These are:

The Learner Belief Scale (LBS); The User Interface Scale (UIS); Content Quality Scale (CQS); Socialisation Quality Scale (SQS); Behavioural Intention Scale (BIS); Perceived Motivation Scale (PMS); System Usability Scale (SUS).

Both the ResQue model and the user-centric evaluation framework of Knijnenburg et al (2012) offer a model to determine the concepts and measures relevant for quality in recommendation. They can guide learning resource platform providers in developing a sustainable evaluation strategy for their platform innovations.

Each of the concepts in the two models can be translated into measures that are relevant for your platform. They can also be modified to the specific design of your implementation.

Conclusions

In this section, we looked at a methodology to implement recommendation engines on learning resource platforms. Here, we give an overview of the main conclusions.

1. Start with a plan: it is important to determine the current user experience and the desired user experience, including the role that recommendation should play in that user experience.

2. Diversification of the recommendations have a positive effect on user satisfaction. However, the extent to which the recommended list is diverse differs from user to user.

3. Very personalized lists, carrying niche items are not appreciated by users (Jannach, D., Lerche, L., & Jugovac, M. (2015).

4. Unfamiliar items also are deemed an obstacle for users working with the system (Jannach, D., Lerche, L., & Jugovac, M. (2015).

5. Graph-based recommenders do seem to have positive effects on user satisfaction.

6. When users are looking for something specific, diverse lists do not work effectively. In these cases, it may be more useful to redirect the user to a good search functionality.

7. The way recommended results are presented to the user has a great effect on the user experience. Arana-Llanes, Rendón-Miranda, González-Serna & Alejandres-Sánchez (2014) even suggest this represents 50% of the user experience. Karga & Satratzemi (2019) suggest that all findings of user-centric evaluations need to take into account the user interface through which recommendations were presented.

8. Giving users control over the parameters in a recommendation engine allows them to change settings and understand recommendations for their specific situation. For many users, this will allow for more user interaction and ultimately more user engagement.

9. It is important to include a dedicated user evaluation strategy, together with any introduction of a recommendation system. This will give insight into how the recommendation engine performs, if and how it engages with users and what can be improved.

Challenges and recommendations

In this section, the primary challenges for the use of recommender systems in learning resource platforms aimed at K12 teachers will be identified. Some recommendations will also be given for each of the challenges.

1. Learning resources are complex items, which can include multiple types of resources.

Learning resources often consist of textual, audiovisual or multimedia data. They can also consist of working formats or games. As such, learning resources are complex entities. Much of the existing research on recommender systems focus on particular types of data. For learning resources, it might require giving attention to all data types as well as interaction between data types.

2. User-centric evaluation of recommender systems is highly influenced by a particular setting and design.

Recommender systems are very much determined by their specific setting, design and use (Knijnenburg et al, 2012). Therefore, from a business perspective, user-centric evaluation needs to be part of the business process around the recommender system. This report presents a way to do install a continuous sustainable user-centric evaluation effectively.

3. Users do not seem to be very aware of differences in technique, or only up to a point. It will therefore pay to not invest all the efforts on getting the "best" algorithm.

Although recommender algorithms are important to achieve good results for a particular user, users themselves are not able to distinguish the minor differences between algorithms when they are using the platform. The primary explanation for this is that recommender systems create the world or platform as the user sees or experiences it. The user cannot step out of this specific, personalized world to distinguish if the recommender system presents the most relevant information that they can receive. From a business perspective, it is therefore more interesting in following the user experience and setting up sustained user interaction with their users.

4. As algorithms are just one part of a recommender system, some effort can be put in the presentation and interaction with the user – on the level of the user interface.

As the previous point unveiled, algorithms are not perceived by users to make a difference. However, giving the user more insight into differences between recommendation strategies and parameters in ranking, will engage the user more in understanding the platform. Giving the user agency here will also create a more complete user experience, which ultimately may lead to more loyal custom and valued use of the platform.

5. Understand the world of your users: they will access your platform at specific moments, on specific devices and with specific purposes.

It is important to have insight into how users use the learning resource platform. This requires insight into the working practices of your target users. Understanding how they approach the platform, when they use the platform, and in which ways will allow you to design your platform better and in a more relevant way. The best user-centric business process to support this is the establishment of a user focus group of core loyal friendly users, who can give continuous in-depth feedback on the user experience.

.

6. Different devices (desktop, mobile) allow for different sorts of interaction with the learning resource platform and content.

Recommender systems have the potential to be very aware of the context in which the user accesses the platform. Understanding how the different devices are used and in which contexts will allow to define better recommendation strategies and better user experiences.

Recommender systems can be a good business decision. As any innovation in strategy, they require good forethought in decision, preparation in implementation and a proper follow-up for the technology to grow into a valuable part of your learning resource platform.

References

Champiri, Z. D., Mujtaba, G., Salim, S. S., & Chong, C. Y. (2019, January). User Experience and Recommender Systems. In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-5). IEEE.

Dias, A. D. S., & Wives, L. K. (2019). Recommender system for learning objects based in the fusion of social signals, interests, and preferences of learner users in ubiquitous e-learning systems. Personal and Ubiquitous Computing, 1-20.

De Pessemier, T., Courtois, C., Vanhecke, K., Van Damme, K., Martens, L., & De Marez, L. (2016). A user-centric evaluation of context-aware recommendations for a mobile news service. Multimedia Tools and Applications, 75(6), 3323-3351.

Fazeli, S., Drachsler, H., Bitter-Rijpkema, M., Brouns, F., van der Vegt, W., & Sloep, P. B. (2018). User-Centric Evaluation of Recommender Systems in Social Learning Platforms: Accuracy is Just the Tip of the Iceberg. IEEE Transactions on Learning Technologies, 11(3), 294-306.

Epifania, F., & Porrini, R. (2016, April). Evaluation of Requirements Collection Strategies for a Constraint-based Recommender System in a Social e-Learning Platform. In Proceedings of the 8th International Conference on Computer Supported Education (pp. 376-382). SCITEPRESS-Science and Technology Publications, Lda. https://dl.acm.org/citation.cfm?id=3096384

Etemadpour, Ronak G., Christopher Crick, Lars Linsen, Jose Gustavo Paiva, and Angus Graeme Forbes. "Etemadpour, R., Linsen, L., Paiva, J. G., Crick, C., & Forbes, A. G. (2015, March). Choosing visualization techniques for multidimensional data projection tasks: A guideline with examples. In International joint conference on computer vision, imaging and computer graphics (pp. 166-186). Springer, Cham. Examples." Communications in Computer and Information Science 598 (2016): 166-86. Web.

Ferwerda, B., Graus, M. P., Vall, A., Tkalcic, M., & Schedl, M. (2017, April). How item discovery enabled by diversity leads to increased recommendation list attractiveness. In Proceedings of the Symposium on Applied Computing (pp. 1693-1696). ACM.

Ferwerda, B., Graus, M. P., Vall, A., Tkalcic, M., & Schedl, M. (2016, August). The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists. In EMPIRE@ RecSys (pp. 43-47).

Jannach, D., Lerche, L., & Jugovac, M. (2015). Item Familiarity Effects in User-Centric Evaluations of Recommender Systems. In RecSys Posters.

Karga, S., & Satratzemi, M. (2019). Using explanations for recommender systems in learning design settings to enhance teachers' acceptance and perceived experience. Education and Information Technologies, 1-22.

Knijnenburg, Bart, Martijn Willemsen, and Alfred Kobsa. "Knijnenburg, B. P., Willemsen, M. C., & Kobsa, A. (2011, October). A pragmatic procedure to support the user-centric evaluation of recommender systems. In Proceedings of the fifth ACM conference on Recommender systems (pp. 321-324). ACM." Proceedings of the Fifth ACM Conference on Recommender Systems (2011): 321-24. Web.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction, 22(4-5), 441-504.

Knijnenburg, B. P. (2016, March). Evaluating Intelligent User Interfaces with User Experiments. In Companion Publication of the 21st International Conference on Intelligent User Interfaces (pp. 6-8). ACM.

Nariman, D. (2013, July). Analyzing text-based user feedback in e-Government services using topic models. In 2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems (pp. 720-725). IEEE.

Ochirbat, A., Shih, T. K., Chootong, C., Sommool, W., Gunarathne, W. K. T. M., Wang, H. H., & Ma, Z. H. (2018). Hybrid occupation recommendation for adolescents on interest, profile, and behavior. Telematics and Informatics, 35(3), 534-550.

Parra, and Brusilovsky. "Parra, D., & Brusilovsky, P. (2015). User-controllable personalization: A case study with SetFusion. International Journal of Human-Computer Studies, 78, 43-67." International Journal of Human - Computer Studies 78.C (2015): 43-67. Web.

Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In Proceedings of the fifth ACM conference on Recommender systems (pp. 157-164). ACM.

Said, A., Jain, B. J., Lommatzsch, A., & Albayrak, S. (2012, August). Correlating perception-oriented aspects in user-centric recommender system evaluation. In Proceedings of the 4th Information Interaction in Context Symposium (pp. 294-297). ACM.

Said, A., De Luca, E. W., Kille, B., Jain, B., Micus, I., & Albayrak, S. (2012, February). Kmule: a framework for user-based comparison of recommender algorithms. In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (pp. 323-324). ACM.International Conference on Intelligent User Interfaces (2012): 323-24. Web.

Shi, L. (2014). Defining and evaluating learner experience for social adaptive e-learning. In 2014 Imperial College Computing Student Workshop. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Willemsen, M. C., Graus, M. P., & Knijnenburg, B. P. (2016). Understanding the role of latent feature diversification on choice difficulty and satisfaction. User Modeling and User-Adapted Interaction, 26(4), 347-389.

Zhao, Q., Harper, F. M., Adomavicius, G., & Konstan, J. A. (2018, April). Zhao, Q., Harper, F. M., Adomavicius, G., & Konstan, J. A. (2018, April). Explicit or implicit feedback? engagement or satisfaction?: a field experiment on machine-learning-based recommender systems. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 1331-1340). ACM. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 1331-1340). ACM.

Zolaktaf, Z., AlOmeir, O., & Pottinger, R. (2018, July). Bridging the Gap Between User-centric and Offline Evaluation of Personalized Recommendation Systems. In Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (pp. 183-186). ACM.