Recommender Systems for Health Informatics: State-of-the-Art and Future Perspectives

André Calero Valdez[1,4(✉)], Martina Ziefle[1], Katrien Verbert[2],
Alexander Felfernig[3], and Andreas Holzinger[4]

[1] Human-Computer Interaction Center, RWTH-Aachen University,
Campus Boulevard 57, 52074 Aachen, Germany
{calero-valdez,ziefle}@comm.rwth-aachen.de
[2] KU Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium
katrien.verbert@cs.kuleuven.be
[3] Institute for Software Technology, Graz University of Technology,
8010 Graz, Austria
alexander.felfernig@ist.tugraz.at
[4] Holzinger Group HCI-KDD, Institute for Medical Informatics,
Statistics & Documentation, Medical University Graz, 8036 Graz, Austria
a.holzinger@hci-kdd.org

Abstract. Recommender systems are a classical example for machine learning applications, however, they have not yet been used extensively in health informatics and medical scenarios. We argue that this is due to the specifics of benchmarking criteria in medical scenarios and the multitude of drastically differing end-user groups and the enormous context-complexity of the medical domain. Here both risk perceptions towards data security and privacy as well as trust in safe technical systems play a central and specific role, particularly in the clinical context. These aspects dominate acceptance of such systems. By using a Doctor-in-the-Loop approach some of these difficulties could be mitigated by combining both human expertise with computer efficiency. We provide a three-part research framework to access health recommender systems, suggesting to incorporate domain understanding, evaluation and specific methodology into the development process.

Keywords: Health recommender systems · Human-computer interaction · Evaluation framework · Uncertainty · Trust · Risk · Privacy

1 Introduction and Motivation

What should I read next? [1] What should I watch next? What product will I find interesting? These are typical questions that traditional recommender systems are designed for. Recommender systems help sieve through large amounts of data in determining options that are most relevant to the task the user has in mind. If the user is a customer using an online retailing platform, recommender systems are pervasive. Almost everyone has seen Amazon's "Other customers have also

bought XYZ" suggesting other products that are relevant to the current search request. Research in recommender systems in e-commerce focus on algorithms, interaction patterns and evaluation. Typical objects of applications are movies, music, documents, or products.

While the choices made by customers can be described as being low-risk, decisions made in other domains may have more severe consequences for the end user. In particular in the area of health and medicine, the limiting resource is the (possibly) non-replenish-able health of the patient. The recommender system should not only avoid failures and support decision making, but it should also understand the patient, the attitudes, the requirements, the values in the context of disease and health management. This makes the applicability of health recommender systems more tricky.

First, we must clarify where recommender systems are applied in the health domain? What are the options to be recommended? Does the system offer therapy suggestions to a doctor or do we supply nutrition-based food recommendations? Both systems are drastically different, yet share inherently similar risks, either for the individual, for the society as a whole, or both.

No framework exists that unites the specificity of health related recommender systems in order to provide both guidance to develop, and metrics to evaluate a health recommender system. In this article we aim to provide a review of recommender systems, how they have been applied in health scenarios, and how we think a framework can help in creating better health recommender systems.

2   Glossary and Key Terms

In order to make this article more understandable for researchers that have no prior experience in recommender systems. We first provide a glossary and some key terms that are relevant in this article.

2.1   General Terminology

Patients are typically persons that are in some kind of medical care. In health recommender systems patients (and in some cases users) must not necessarily be suffering from an illness. Health recommender systems may largely be applied as preventive measures, as well.

Machine Learning (ML) addresses the question of how to design algorithms that improve automatically through experience [2] - the focus is on doing it automatically (aML) without a human-in-the-loop [3].

interactive Machine Learning (iML) can be defined as learning algorithms that can interact with both computational agents and human agents and can optimize their learning behaviour through these interactions [4], by bringing in a human-in-the-loop [5].

2.2   Recommender Systems Terminology

Users – People that use the system and receive recommendations. Users also provide the ratings for items.

Items – The items that are being recommended (e.g. movies, products, hotels, etc.).

Ratings – Ratings refer to the choices of users in relation to items. Ratings can be explicit, by e.g. tagging a product, or implicit (e.g. opening a document, buying a product). Ratings can be Boolean, ordinal or numeric in nature, requiring different algorithms in implementation.

Content – The data from within the items that can be analyzed for recommendation. When documents are recommended, typically the content of the document. In many cases also meta-data.

Task – The reason why a user uses the system (e.g. to find a movie to watch). Often a set of interdependent tasks are relevant.

Context – The sum of all contextual factors that influence the use and evaluation of the recommender system and their interactions.

Sparse Matrix – A matrix that contains mostly the value 0. In user-item matrices, we often have many users and many items and only few ratings for individual users and items.

Cold start problem – When a new item or new user enters the system, we have very little information on the user to base recommendations on.

Coverage – Coverage refers to the criterion that addresses, whether all items in the database are getting recommended [6]. Recommenders that only recommend the most bought items, reach low coverage.

Serendipty – Serendipity refers to the criterion that addresses, whether recommended items are unexpected to the user. Novel unexpected items — serendipitous finds — can be a core benefits of a recommender.

3   Recommender Systems

The purpose of a recommender system is to find items that are relevant to the user, based on the users previous decisions. Recommender systems use these decisions and the decisions of other users to establish what may be relevant to the user. The first "recommender system" can be traced back to the Tapestry system developed at XEROX Park [7]. The initial idea was the concept of *collaborative filtering*. Users would tag interesting items to allow other users to browse these items by tags. The principles behind collaborative filtering are still being used in recommender systems [8].

Research then quickly focused on identifying similarities within documents. This allowed to rely not only on the opaque choices of users but also on the

content of the recommended items, hence the name *content-based filtering* [9]. Quickly hybrid approaches appeared, merging both content-based and collaborative filtering [10].

Modern recommender systems often use *compositional approaches*, combining multiple recommender algorithms to an overall solution [11]. Techniques from other fields of computer science also find their way into recommender systems research. For example, social network analysis is used to augment recommendations [12] with data from relevant peers. Here, research on trust-based recommendation [13,14] has shown, that recommendations given by trusted peers are more likely to be helpful than generic algorithmic solutions. Recent approaches have also used methods like deep learning [15,16] to uncover the non-linear structure of preference in users.

Overall, we can say that research focuses on algorithms, data sets, evaluation criteria and interfaces for recommender systems. To each of those areas we provide a short introduction in the following sections. We then see, what of this work has been applied to health recommender systems, and what still needs work.

3.1   Algorithms

One typical research area, also the one with the strongest focus, is the underlying motor of the recommender system – the recommendation algorithm. Several approaches are used depending on the context of use. This sections aims to give a quick overview over the field of algorithms in recommender systems. It is by no-means extensive, but aims to help the reader to understand later parts of this article where algorithms are used.

Collaborative Filtering relies on the individual ratings of all users. It tries to identify items that are relevant for the user, or related to other items that have received positive ratings by the user. For this purpose a user-item matrix is used (see Fig. 1 on the left). This matrix simply contains the ratings of the users for all items. When a user has not rated an item, the cell remains empty. Typically this is a sparse-matrix. Various methods are used to impute the empty cells. This imputation is a prediction of a rating from a user for a given item. Predictions of high ratings can be used as recommendations. Some of these methods are given here:

– *Row mean* – By utilizing the mean of the row we average the users rating and return an non-informative rating, matching the users average judgment.
– *Column Mean* – By using the mean of the column we utilize how users have rated a particular item on average.
– *Combined* – We can use both means to adjust the prediction of an item to the respective users rating behavior.
– *Row based cosine similarity* - The vector cosine allows comparing the similarity of two vectors. The cosine is 1 if vectors are the same and 0 if they are
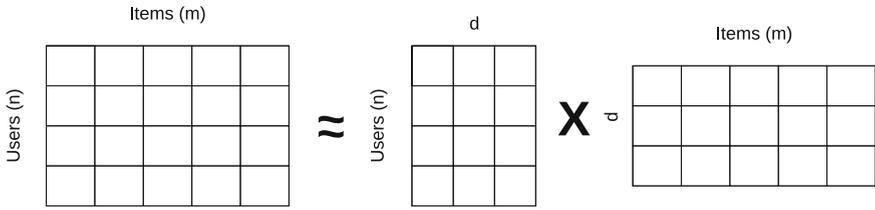
Fig. 1. Non-negative matrix factorization. By factorizing the user-item-matrix on the left, we can extract latent commonalities (d) of both users and items and calculate recommendations faster.

orthogonal. We can use this to find similar users and use their rating as a means of prediction.

– *Column based cosine similarity* – This can be used to find similar items and using them for recommendations immediately.

In order to improve run-time performance and to overcome the sparse-matrix and a part of the cold-start problem, we can use *matrix-factorization* to determine latent commonalities (here d) of both users and items (see Fig. 1). Matrix factorization has been found to be superior to nearest neighbor procedures and allows to integrate additional information (implicit feedback, temporal information, etc.) [17].

The natural extension of matrix factorization is *tensor-decomposition*. When information on ratings contains a third dimension of information (e.g. location-data, social preferences [18], context [19], etc.), we can encode this information in additional tensor dimensions. In order to apply similar procedures, we can no longer rely on matrix factorization, but must use tensor decomposition to compute recommendations [20].

Depending on the runtime-complexity of calculating latent preferences, different approaches exist, when incorporating new ratings. If complexity is high, different recommendation techniques can be combined to address users who are new to the system and thus not adequately represented. This can be done until the latent preferences have been updated.

Content-Based Filtering uses meta-data or features from individual items to open the black-box of the non-descriptive "item-id". In document recommendation typically text-mining methods are used for feature extraction. A typical text-mining pipeline would include the following steps, and yields a *vector-space model*:

1. *Term-Document Matrix* – Used to store a bag of words model of all documents.
2. *Stop-Word Deletion* – Used to remove words that are not predictive (e.g. "this")
3. *Stemming* – Using the stem of a word only (e.g. "walk" instead of "walking").

4. *N-Gram Detection* – Finding words that appear frequently together which are used as a singular term (e.g. "recommender system").
5. *TF-IDF* – Used to weight words in accordance with their relative importance for the document at hand.
6. *Latent-Semantic Indexing* – Using singular value decomposition on the term-document-matrix to incorporate semantic information in the extended vector space model.

As the end result we have three matrices that can be used to compute similar documents, based on the semantic similarity of documents. As an alternative approach we could use *Latent-Dirichlet Allocation* in order to identify topics in documents and find documents with similar topic-distributions. These similarity measures can be used in a similar fashion as in the collaborative filtering. When the recommended items are not documents, one must consider what are features of items that are relevant for recommendations (e.g. product features, actors in movies, etc.).

3.2   Data Sets

In 2006 Netflix[1] released a part of their users' ratings data sets. They proposed a challenge open for anyone, to train an algorithm to outcompete their implementation in predicting ratings of the remaining (unpublished) data. The *Netflix price* spiked the development in recommender algorithms and yielded Bellkor's Pragmatic Chaos algorithm [21]. The algorithms were measured against the root-mean square error (RMSE) of their predictions with the actual data. Interestingly, this algorithm was never included in the Netflix system. Partly because the algorithm was optimizing against an irrelevant metric. In 2006, Netflix believed that by reducing the RMSE, better recommendations could be achieved. But the RMSE can also be reduced by optimally predicting the ratings of movies, which in the lowest interest of the users. This might reduce RMSE but helps only little with good recommendations.

Other data sets exist from bibsonomy (scientific publications), delicious (bookmarked hyperlinks), flixster, movielens, movietweetings (all on movies), million songs dataset (music), and ta-feng (grocercy shopping bags from Belgium). Further data sets are described at the RecSysWiki [2].

3.3   Evaluation of Recommender Systems

As we have seen before, the criterion, against which a recommender system is evaluated, is critical to its success. Traditionally, recommender system algorithms were evaluated based on criteria borrowed from information retrieval or signal detection theory. Typical metrics are [22]:

---

[1] Netflix is an online movie provider.
[2] http://www.recsyswiki.com/wiki/Category:Dataset.

– *Precision* – The percentage of relevant items that are correctly recommended out of all recommended items.
– *Recall* – The fraction of items that are recommended from all relevant items. Also called *sensitivity*.
– *F-Measure* – The F-Measure is the harmonic mean of both precision and recall. It combines both measures into a single metric (see also Fig. 2).
– *ROC-Curve* – The reciever operating characteristic is a plot that visualizes the change of true positives against false positives depending on the sensitivity threshold. A sensitivity threshold must be chosen, as typical output of algorithms is mostly never clearly 1 (recommend) or 0 (do not recommend). The ROC-Curve helps determining optimal thresholds and compare algorithms against each other independently of the selected threshold.
– *RSME* – The root-square mean error is a measure that can be used to compare predictions against real data. By calculating the squared error for all items and then taking the root of the mean of the squared errors, we receive a value that penalizes strong deviations and is relatively forgiving to small deviations. This yields a weighted score which increases, when predictions differ more strongly from real values, and decreases when predictions become more accurate.

But as the Netflix price showed, that reducing error alone does not help in creating a better recommender. Ge et al. therefore argued to move beyond recommendation accuracy [23]. In a recommendation scenario often only the first k items are viewed by the user. If none of these are relevant, a user might go to a different website and not buy anything, in an e-commerce setting. This idea yields the k-top recommendations metrics (i.e. how many of the k-top items does the algorithm correctly find for all users?). But also measures such as serendipity (i.e. are items new?) and coverage (i.e. are all items being recommended?) are important, because users do not want to see their most beloved movies over and over again.
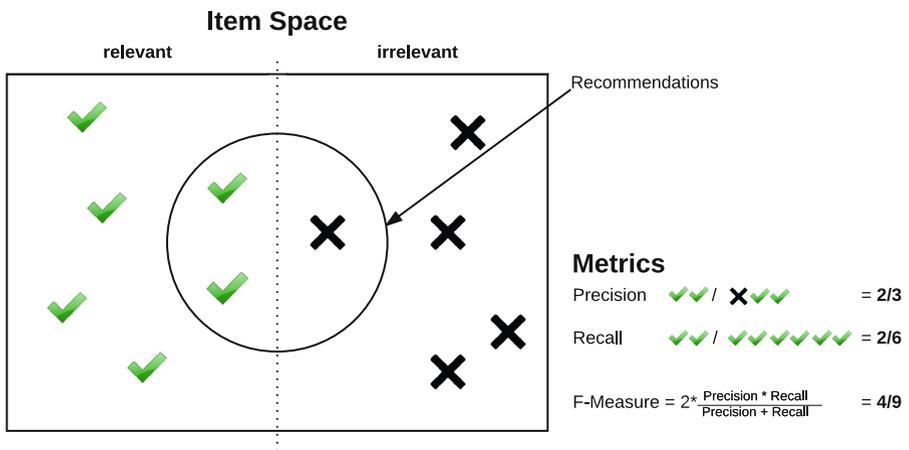

Fig. 2. Visualization of three metrics in a recommendation example: precision, recall and F-measure

3.4   Interfaces and Interactive Recommender Systems

When considering the whole recommender system in real usage scenarios, it became clearer that not only the algorithm needs to be evaluated [24]. The interface and HCI of a recommender system are equally important in how a user interacts with a recommender system.

This led to the design of user-centered evaluation frameworks. Most famous the work by Knijenburg et al. [25] and Pu et al. [26]. The work of Knijenburg et al. suggests that depending on **domain knowledge**, different types of interactions are most helpful to the user. Novices and "maximizers" prefer top-recommendations, while experts prefer hybrid-approaches combining implicit and explicit preferences.

Pu et al. shifted the focus of evaluation on technology acceptance. The new criteria were categorized as perceived quality, user believes, user attitudes and behavioral intentions. Under **user perceived quality** we summarize the perceived quality of recommendations (e.g. accuracy, familiarity, novelty, etc.), the interaction adequacy (e.g. adequacy of expression of ratings, explanations, etc.), and the interface adequacy (e.g. layout adequacy, clarity, etc.). The perceived quality then influences the **users' beliefs** about the recommender system. User beliefs concern the perceived ease of use, perceived usefulness, and control and transparency of the system. The **users' attitudes** refer to the attitude the user has in regard to the recommender systems. These encompass attitudes such as overall satisfaction, confidence and trust in the recommendations. Lastly different types of **behavioral intentions** can be measured on the base of cognitive and motivational attitudes of users. The user can either be willing to use the system, buy a product, continue to use the system or even influence his social circle to use the system.

As the evaluation of recommender systems took a turn to the user, of course the interface to the recommender became more important. In their review on interactive recommender systems, He et al. [27] reveal that interactive recommender systems aim at fulfilling the evaluation criteria *transparency*. By visualizing items in non-list based manners, and by showing how a recommendations come together, the results are more transparent to the users. Some of these interfaces are designed to foster exploration [1] and serendipity, others to provide overview and explanations [28].

Although the HCI part of recommender systems has become increasingly important, it still takes a smaller part in recommender research overall [29]. In particular, aspects such as user-control, affective interfaces and high risk domains — such as health — have not had a large share of research.

3.5   Health Recommender Systems

Not much previous work on applying recommender system in health informatics or medicine exists. As of June 5th 2016 only 17 articles are found when searching for the terms "recommender system health" in web of science. The oldest article is from 2007 and the most cited article has only 14 citations. Wiesner and Pfeifer

[30] distinguish between two scenarios: the first scenario targets health professionals as end-users of health recommender systems. The second scenario targets patients as end-users. Health professionals can benefit from recommender systems to retrieve additional information for a certain case, such as related clinical guidelines or research articles. The second scenario focuses on delivering high quality, evidence based, health related content to end-users. Most other articles that we have reviewed target patients as end-users. Objectives include delivering relevant information to end-users that is trustworthy, as in the work of Wiesner and Pfeifer [30], lifestyle change recommendations [31] and improving patient safety [32]. The latter category for instance includes research on how to use recommender systems to suggest relevant information about interactions between different drugs, in order to avoid health risks. Lifestyle change recommendations focus among others on suggesting users how to improve their eating [33,34], exercising or sleeping behavior.

In their research statement Fernandez-Luque et al. [35] argue, that using recommender systems for personalized health education does not take advantage of the increasing amount of educational resources available freely on the web. As one reason, difficulties in finding and matching content is given.

In a short review on health recommender systems by Sezgin & Ozkan [36] provided at the EHB 2013, the authors emphasize the increasing importance of Health Recommender Systems (HRS). The authors argue, that these systems are complementary tools used to aid decision making processes in all health care services. These systems show a potential to improve the usability of health care devices by reducing the information overload generated from medical devices and software and thus improve their acceptance.

The 2016 ACM Conference on Recommender Systems conference featured a workshop on engendering health with recommender systems, where many of the topics from this article were discussed.

4   A Framework for Health Recommender Systems

In order to successfully develop a health recommender system, additional criteria and procedures must be incorporated to ensure the success of such a system. The area of health or medical recommender systems faces several challenges that make it specific and intricate.

First of all, there is no clear task definition for recommender systems in health. The purpose of a recommender system depends on the item being recommend. In a health scenarios various items are imaginable. For example, a rather typical recommender system in a mobile device could recommend physical activities that match the current user situation to improve their health. An patient with arthritis and obesity could benefit more from physical activity recommendations that put no additional strain on already inflamed joints. Going for a walk will be more pleasurable, if weather conditions are good. Another, very different example of a health recommender could be a system that proposes different forms of cancer therapies to both a doctor and the patient. The
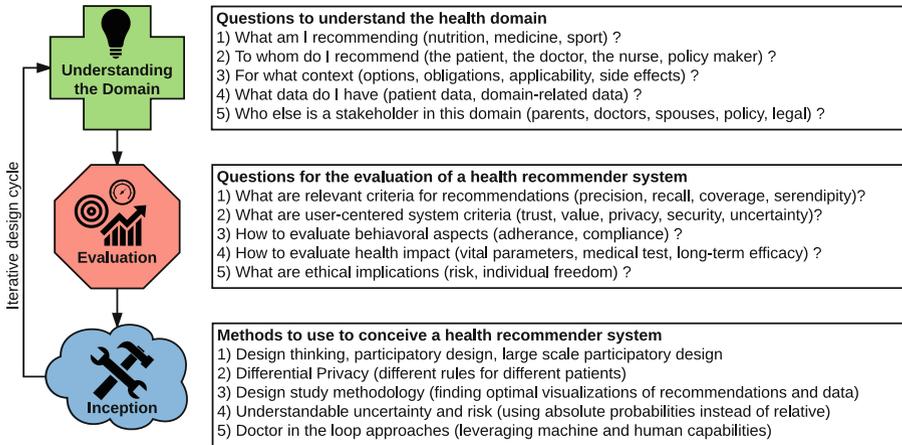
Questions to understand the health domain
1) What am I recommending (nutrition, medicine, sport) ?
2) To whom do I recommend (the patient, the doctor, the nurse, policy maker) ?
3) For what context (options, obligations, applicability, side effects) ?
4) What data do I have (patient data, domain-related data) ?
5) Who else is a stakeholder in this domain (parents, doctors, spouses, policy, legal) ?

Questions for the evaluation of a health recommender system
1) What are relevant criteria for recommendations (precision, recall, coverage, serendipity)?
2) What are user-centered system criteria (trust, value, privacy, security, uncertainty)?
3) How to evaluate behiavoral aspects (adherance, compliance) ?
4) How to evaluate health impact (vital parameters, medical test, long-term efficacy) ?
5) What are ethical implications (risk, individual freedom) ?

Methods to use to conceive a health recommender system
1) Design thinking, participatory design, large scale participatory design
2) Differential Privacy (different rules for different patients)
3) Design study methodology (finding optimal visualizations of recommendations and data)
4) Understandable uncertainty and risk (using absolute probabilities instead of relative)
5) Doctor in the loop approaches (leveraging machine and human capabilities)

Fig. 3. Three steps to consider when developing a health recommender system. These steps should be incorporated as an extension of typical steps.

system could integrate patient properties such as other illnesses, additional medication, job requirements, and family situation to recommend optimal therapies and alternatives. It could visualize duration, experience and possible side-effects of multiple therapies and thus increase the patients control over their situation. The recommender system could be a communication tool that is used by both doctor and patient to help make difficult decisions.

In both scenarios the underlying algorithms can be taken from recommender systems research, but serve drastically different purposes and thus change the requirements to the recommender system.

In order to help understand the design space of such recommender systems, we propose the use of three additional design steps (see Fig. 3) when conceiving a health recommender system. Each step proposes guidance questions or additional methods and procedures to enrich the contextual picture of the usage scenario. We propose to extend the traditional recommender system design procedure to encompass theses additional requirements.

4.1   Understanding the Domain

First, we believe that different questions are necessary in understanding the application domain. As with any other recommendation domain, we must first understand what the **recommended item** is. Possible categories are:

– *Food/Nutritional Information* – Providing recommendations to optimize nutrition. May be applied to compensate malnutrition, reduce weight or to prevent certain food-based illnesses [33,34]. Recommendations could be food replacement items [37,38], different meals [39], or additions to a diet. The complex nature of taste [40] and its temporal and social dependencies [41] have to be considered.

– *Physical exercise/Sport* – Providing recommendations on what physical activities to perform. May be applied to help in finding activities that are interesting and motivating and also match the users requirements and needs. A recommender could also include location-data and weather data to find activities that are optimal for the users context.
– *Diagnosis* – Providing recommendations on likely diagnoses of a patient to a doctor or nurse. An approach using this recommendation item can be very similar to case-based reasoning approaches. By adding recommendations and linking them to therapy options, further value can be created.
– *Therapy/Medication* – Providing recommendations about the variety of possible options that may be applicable. A recommender system could address either/both patient and professional in finding a patient-specific therapy. The therapy, as mentioned in the example before, could include various patient properties and visualize different outcome criteria. Recommender systems could create personalized-health solely from data-analysis.

The second important question is, who are the **users** for the domain at hand? Typically the system is designed for an end-user, who can be either healthy or already a patient. But health recommender systems may extend their audience to health professionals such as doctors and nurses. Beyond these obvious new stakeholders pharmacists, clinicians, researchers and also policy makers could benefit from health recommender systems. Reducing the cost of health care in general could be a goal of recommender systems.

The third big part of the domain is the **usage context**. The context contains both the multifactorial goal setting, as well as contextual factors that influence how items are recommended and how they should be presented. By *multifactorial goal setting* we mean that health goals are not following a singular dimension. While naively we might think that the "most healthy" option is the one that should be recommended, different domain-specific criteria play a role in evaluating an item. For example, what is healthy for one patient could be dangerous for another (e.g. diuretics and other blood pressure lowering agents are dangerous for patients with diabetes, gout, etc.). It is necessary to include side-effect reduction as a goal. Beyond these immediate health-related outcomes, other outcomes such as costs, applicability (e.g. is the patient able to perform a daily subcutaneous injection) or changes in quality of life are important to consider. The impact of individual goals is also expected to differ strongly between different diseases. One can easily imagine that the relative importance of different goals vary drastically from one another for illnesses such as gout, cancer, depression or allergies. The goals can both be finding alternatives or finding optimal solutions. Thus the recommender system could very easily become a decision support system, depending on the dimensions of the search-space. Some illnesses have very few effective therapies, while other illnesses may have a multitude of tools that still only help to alleviate symptoms. In the latter scenarios complex recommendations (i.e. one therapy per symptom) could be the outcome of a recommender. The overall compatibility with the patients predispositions could be judged as the quality of a recommendation. It is also important to consider that some

patients may value quality of life over longevity. A cancer patient might refuse painful therapy in order to enjoy the last few months at home rather than in a hospital. Optimizing for highest probable health outcome, by comparing efficacy of different medicines, could be an obvious solution but not necessarily the one a patient might have chosen in retrospect.

The fourth area of the domain is the **availability of data**. While typical areas of recommendations — such as movies — have the benefit of having publicly available data-sets, which can be used to train and test algorithms, data sets for health recommendations are rare. Some of the problems stem from the intrinsically more *complex nature of health data*. Health data is often unstructured, incomplete, non-standardized and stems from various sources. Large parts of data are not generated in a computer (as typical recommendations are) but stem from paper-based health-records that are often digitized afterwards. Additionally, the *type of data* may differ. While we have large nutritional databases, understanding what food stuffs serve as tasteful replacements is not fully understood. Also different brands of similar food stuffs make recommendations harder to do. IBM Watsons chef for example recommends dishes using IBM Watson technology. But simply trying to exclude sugar from recipes, requires to exclude 19 different types of sugar.

To make things more complicated health data has inherent *privacy* issues. Non-anonymized patient data in the hand of insurance companies, can be in the disinterest of the patient.

Stakeholders who are relevant besides the user should also be considered. In cases of health recommenders for food, for example, not only the user is affected by the system. Food is (most) often consumed in groups, thus individual preferences of multiple users influence the possibility of choices. Very specific choices or recommendations (e.g. low-carb, low-fat, sugar-free gluten-free, vegan) might be contradictory with each other. In such cases it is thinkable, that the group preference might overturn the recommended solution. Another approach could be to find group recommendations [41]. In another scenario, the parents of a child that suffers from diabetes are core stakeholders. Moreover, any family care givers could also benefit from high quality health recommender systems, as they might be the ones administering the care.

4.2   Evaluation of Health Recommender Systems

The criteria mentioned in this section should be considered as additions to typical evaluation criteria of recommender systems (e.g. [26, 42, 43]).

As with any medical technology it is crucial to measure and benchmark health recommender systems, particularly in regard to user acceptance [44] and satisfaction. By tailoring the services or devices to the individual user needs, demands and requirements, future research issues are uncovered. This includes user diversity research, not just in regard to user tailored results, but also in regard to the user interface of a recommender system [45]. Questionable quality will not be acceptable in a health recommender system.

To enable *benchmarking*, more comprehensive quality measures must be sought, and more specific theoretical and experimental frameworks should be investigated [43]. The overemphasis of accuracy metrics and under-representation of metrics such as serendipity and coverage pose a serious problem in typical recommender systems, but how do they apply to health recommender systems?

In particular rare diseases are important in this regard. Rare disease are individually rare, but there a many different rare diseases, making patients with rare diseases a non-rare phenomenon. Therefore, the question of serendipity and coverage, i.e. finding "interesting" as well as finding all relevant results, is also important for health recommender systems.

Algorithms exist to trade-off **serendipity and coverage** [23] for improved accuracy, e.g. by recommending items with more data. And since all three of these measure are important to users in many applications, designing algorithms in such a fashion may be adequate, but less useful in medical scenarios. Here a DiL approach could also be helpful by integrating the doctor in the algorithm. Judgments that are inherently human [46] (e.g. what is interesting?) can be integrated in the recommendation process, but we will need more comprehensive measures of quality combining accuracy, serendipity and coverage, to allow algorithm designer to improve trade-offs adjusted to medical scenarios in DiL settings.

Another very important research issue is trust in recommender systems [13]. This is particularly true for health recommender systems, as they shall be used to provide end users with more proactive and personalized information relevant to their health. But, there are still many open research questions considering **trust, privacy and intimacy** in the use of medical technology. User diversity plays a role, with an emphasis on gender and age [44].

This is important in regard to user **satisfaction**. Herlocker et al. [43] suggest to look deeper into modeling user satisfaction, with the aim of predictive satisfaction models. In the case of health recommender systems, this prediction is peculiar, as there are different relevant user groups. Differences in expertise, overview knowledge, but also tasks must be understood to create recommender systems suitable to health practitioners, clinical doctors, biomedical researchers, care givers and patients, alike.

As the outcome of health recommendations are inherently uncertain, **communication of this uncertainty** is highly important. Finding ways to visualize uncertainty in a set of recommendations is crucial to allow the user to evaluate the option adequately. This problem is linked to the risk and duration of the consequence of a choice. Picking a bad movie may cost you 90 min of your life; picking a bad therapy could reduce quality of life for many years. This changes how typical evaluation criteria (e.g. k-top recommendations precision) are judged. One bad option in the first few recommendations could have drastic outcomes. The designer of a health recommender systems must be careful and act responsibly in both generating recommendations and communicating them.

Under the assumption that the user of the system has perfect access to the desired options, the effectiveness of such a system must still be evaluated in regard to the users external behavior. **Behavioral evaluations** must be considered to measure the effectiveness of a health recommender system. For example, when giving recommendations about activities to conduct to improve fitness, the recommender system must track what activities have actually been conducted. In the case of smoking cessation [47], the system has a harder time to measure its effectiveness, as users might want to skip reporting that they have smoked a cigarette because of social desirability. Some health recommenders may also aim at long term behavioral changes and these must be tracked somehow, too. The risks of ignoring behavioral changes in long-term evaluations, could lead to short term recommendations, that are helpful to many users in the short-term, but conflict with long term goals (e.g. crash diets) [48].

Measuring actual health impact is also important. Even when the users show long term adherence to recommended health behaviors, the next question is whether the conducted changes in behavior or therapy lead to the desired changes in health. We must consider which health parameters to assess and which medical tests to employ to ensure medical effectiveness. For example, crash diets may lead to reduced body weight (a superficial health parameter), but mostly because of reducing body muscle mass. This leads to rebounding effects because of reduced metabolism. Long-term weight loss is burdened.

Before such an approach can be implemented in real-world medicine, it must be assured that such systems are sufficiently trustworthy [13]. Publicly accessible systems such as collaborative recommenders pose a **security risk**, as normal end-users cannot be distinguished from potential attackers. Beyond these technical risks, such attacks may lead to a continuous degradation of trust in the objectivity and accuracy of such a system. Therefore, a cornerstone of future research is in modeling such attacks and examine their impact on recommending algorithms. Hybrid algorithms in a DiL paradigm could provide a higher degree of robustness to such attacks [49].

Furthermore we must consider **ethical considerations** of recommender systems. (e.g. what do we do when health parameters used as an indicator for disease seem to not correlate with actual disease [50]). The principle of "first do no harm" should be kept in mind [48]. A recommender system might unintentionally provide health guidance that could — in the hand of a person suffering from a mental illness — steer a patient in an unhealthy direction (e.g. dieting tips for anorexic patients).

4.3   Methods to Design Health Recommender Systems

Third we need a framework to help us design a health recommender system in collaboration with the end-users. The aim in designing a recommender system for health should always integrate the end users. A framework needs to integrate tools that bring together the requirements of the domain and the evaluation criteria for the specific application. These tools should help the designer focus on the user and put the user-perspective first.

We think the first tool crucial to success, is the use of **participatory design** [51, 52]. When health is the end-goal of a recommender system, the user should get an active say in designing their system. This helps identifying actual *user needs* and creates *identification* with the future system. When users design the system, the recommender system is not a tool devised by "Big Pharma" to optimize sales, but can become their personal assistant helping them to overcome health burdens that are meaningful to them. It may also alleviate privacy concerns [53]. The challenging part is to extrapolate existing methodology to allow large-scale co-creation and participation. We need tools that allow users to customize their own recommender system and to communicate needs more directly.

In scenarios were users are too remote, or too uninformed to directly cocreate, methods such as *design thinking* [54] can be employed, to ensure that users problems are the focus of the system and not some accuracy metric.

The second tool we think is essential to any health recommender system, is the use of **differential privacy** [55]. Differential privacy allows the sharing of data without revealing individual identities. Typical methods include *k-anonymity or l-diversity* [56], which work by revealing only the amount of data that can be traced back to a group of at least k people. L-diversity additionally integrates the differences in sensitivity of different data fields. Beyond these purely IT-based privacy tools, it is necessary to communicate privacy concerns to the end user. Users are often not aware of privacy risks and behave in manners that contradict their long-term interests. On the other hand many users openly agree with sharing private information, even when aware of risks. Finding the matching *trade-off between privacy and utility* for the individual user group or user is crucial to implementing health recommender systems [57]. In this context, individual factors play a decisive role. As such, the level of knowledge about privacy threats in the Internet is important [58], but also different risk perceptions, as well as the level of digital competency [59], which is often related to age and technology generation [60, 61]. Factual (technical) privacy threat is furthermore not identical with *risk perceptions*. The perceived benefits from sharing the data in the medical context is seen different than sharing the data in a less sensitive field [58, 62, 63]. In addition, users are much more reluctant to share data in personal spaces, when data relate to intimacy contexts, such as homes [64] or the sharing of physiological data [65].

The third tool to incorporate is **adequate uncertainty and risk communication** (e.g. risk-ladders, shaded error-bars, etc.) [29, 66]. Communication goes in both directions – to the user and to the recommender. Users should be able to communicate uncertainty in their input methods, to ensure understanding on the algorithm's side [67]. As the end users may also have differing models of risk and different degrees of understanding statistics and uncertainty, it is crucial to address all levels of risk-literacy (e.g. absolute or natural frequencies [68]).

Furthermore, **effective and efficient visualizations** should be used, when displaying health data. The visualization of data should address the purpose of the recommender system and regard the end user and their intentions.

A methodology to ensure this is the Design Study Methodology [69], which has been shown to be effective even in visual recommender system design [28,70]. Creating a visualization in a recommender fosters the users willingness to explore options and help explain individual recommendations [71]. The challenge is that the influence of user diversity has not yet been fully investigated in information visualization [72]. Since individual differences might play a leveraging role in personalized health applications it is crucial to strengthen research in this areas.

Lastly, assisting medical professionals in a **doctor-in-the-loop (DiL) approach** is a new paradigm in information driven medicine [73]. It pictures the doctor not only as a consumer of digital information, but also as a someone who can interactively manipulate algorithms and tools. The doctor as a final authority inside the loop of an expert system can make sure that expert knowledge is integrated in the decision making process, by finding patterns and supplying tacit knowledge, while the recommender system can integrate patient data as well as treatment results and possible (side-)effects related to previous decisions. The DiL-concept can thus be seen as an extension of use of knowledge discovery for the enhancement of medical treatments together with human expertise: The expert knowledge of the doctor is enriched with additional information and expert know-how [74–76].

5 Conclusion

Recommender systems are applied in almost all fields of commercial web applications helping users to find products and services relevant and interesting to them. They are used to help find interesting information, scientific documents, and collaborators. But even in these areas further research is required [77]. In the future, we will hopefully see health recommender systems integrating experts in the algorithms, thus combining human expertise with computer efficiency to improve medical care for patients, care giver and doctors, providing better health for everyone.

The framework that we have suggested (see Fig. 4) in this article is considered to guide a developer into getting a holistic picture of the constraints that a medical application gives. A medical application is often judged against all of these (and many more) criteria, and inadequately addressing one of these aspects ensures failure of the recommender system and loss of trust in the recommender systems for health in general.

It is of **utmost importance** to mention that this framework is not an extensive one. It does not address any aspects of law, policy, and medicine directly. Developers should take an *interdisciplinary approach* seriously when designing a health recommender system. Seek out professionals from these field, when considering developing a health recommender system. Our framework merely looks at the challenge from a HCI perspective extending into the areas of communication, information visualization and technology acceptance.
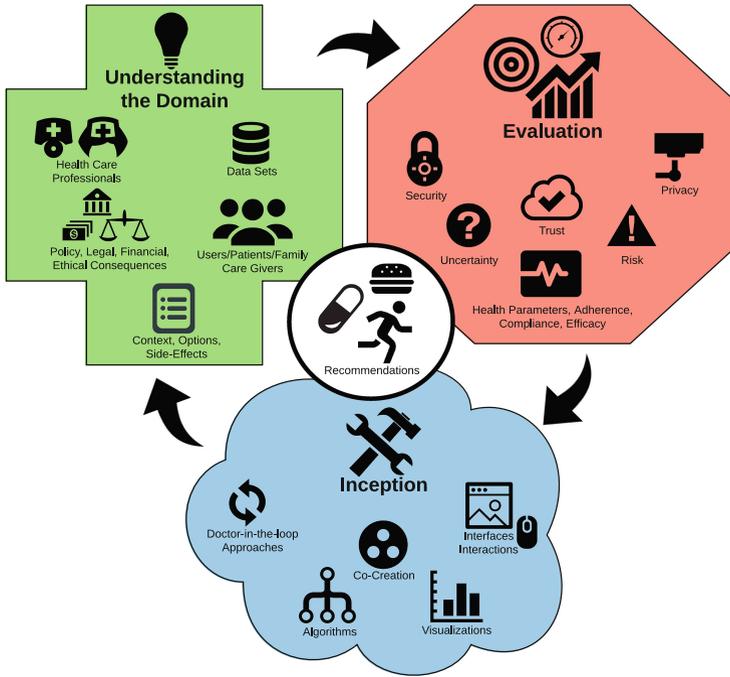
Fig. 4. The three parts of our research framework for developing health recommender systems.

6   Future Work

The future of health recommender systems strongly relies on interdisciplinary collaboration and collaboration across organizational borders. Recommender systems have started to flourish when data sets become public and quality metrics became available. We hope to see more **open data sets** for health recommender system that are helpful in designing algorithms, testing user experience and developing new metrics for the field of health recommender systems. Crucial in this regard is to respect privacy and anonymity, by some means provided earlier in this article. Next to offline evaluations possible through these data sets, it is worth noting that online evaluations still play an extensive role in evaluating a recommender system. User actual reactions might differ drastically from predictions made from offline data [78].

How different types of **user-diversity** (e.g. personality) can be used to improve recommendations has not been fully explored [79]. Even more so for the field of health recommender systems. As one long-term trend of recommender systems could be their integration as a personal assistant —think Siri or Cortana— it will become necessary to think about what health assistant users will want to have [80]. While some users might prefer a purely informative style of assistance others might want an assistant that is responsible for their decisions.

Can a recommender system be responsible for its recommendations? Can the developer be held responsible?

When health recommender systems have been more established, new metrics can be designed to economize recommendations. Sharing data is a form a giving monetary value and receiving helpful recommendations has financial value as well. We think that research should also address this aspect of hidden transactions and develop metrics to measure the **price and value of recommendations** not just from a user perspective, but also from a societal perspective. We are *not* arguing that users should think of recommendations as a product, but any viable recommender system will be used commercially and the consequences will have monetary effects on end users. Understanding the intricacies of how this aspect influences usage of a recommender system is important and should be included in the system design. Models from game-theory could be considered to depict these processes and develop business-models that help foster health and not a single pharmaceutical manufacturer, for instance.

Beyond this economical considerations, research is needed to **evaluate societal impact**. Neither all individual nor all societies can afford the best therapies for themselves. How will health recommender systems address this aspect of applicability and affordability? Maybe we even reduce effectiveness of a treatment, just by showing therapies that are too expensive for a patient [81]. Therefore a naive recommender system could —if globally applied— deteriorate overall health. The authority of the doctor also lies in matching medical needs and financial possibilities. If recommender systems come into play how does this affect a society as a whole? Does it affect the trust in medical professionals? Does it raise distrust in elites?

This also brings up the difference in **culture** and the applicability of recommender systems. In particular when looking at food or nutritional recommendations it is necessary to incorporate effects of culture and cultural tastes [82]. Beyond this superficial limitations, differences in health perception may play a deeper role, when building a health recommender system. In particular, the cultural differences in perceptions of gender or ethnicity may play into the design [83]. Questions that need be raised go in the direction of: How is technology perceived within different cultures? What is the effect of culture on perceptions of risk, privacy, and uncertainty? Does culture play a role in determining the role of individual and group benefit?

Lastly, the area of **ethical implications** of health recommender systems must be explored. While we do know some effects of traditional recommender systems, such as the filter bubble [84], we have not fully figured out the long-term consequences of these effects. When applying these effects to the area of health, the risk of overlooking relevant options might be much more costly for an individual of for society as a whole. But beyond these transferable effects, we must consider further ethical implications.

When recommender systems for health become effective and help in reducing health care cost, the question may become disconcerting whether individuals are still allowed to withhold their data because of privacy concerns. How much

individual freedom is worth how much global health expenditure. The unwillingness to share medical data might increase the cost of therapy and therefore prevent funds to be used elsewhere, indirectly costing other people's health and lives. When recommender systems can be benchmarked in a fashion that makes this cost tangible, will they effectively kill privacy? Does revealing the space of possibilities (and thus the space of impossibilities) help improve health for everyone or only for a selected few? How will pervasive personalized health recommendations influence individual psycho-social development and in extension the Zeitgeist? We, as a society and as researchers, must find ways to decide what role recommender systems will play in the future — both in health and in other fields.

References

1. Bruns, S., Valdez, A.C., Greven, C., Ziefle, M., Schroeder, U.: What should i read next? a personalized visual publication recommender system. In: Yamamoto, S. (ed.) HCI 2015. LNCS, vol. 9173, pp. 89–100. Springer, Heidelberg (2015). doi:10. 1007/978-3-319-20618-9_9

2. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)

3. Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N.: Taking the human out of the loop: a review of bayesian optimization. Proc. IEEE **104**(1), 148–175 (2016)

4. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Springer Brain Inform. (BRIN) **3**(2), 119–131 (2016)

5. Holzinger, A., Plass, M., Holzinger, K., Crişan, G.C., Pintea, C.-M., Palade, V.: Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-ARES 2016. LNCS, vol. 9817, pp. 81–95. Springer, Heidelberg (2016). doi:10.1007/978-3-319-45507-5_6

6. Puthiya Parambath, S.A., Usunier, N., Grandvalet, Y.: A coverage-based approach to recommendation diversity on similarity graph. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 15–22. ACM (2016)

7. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM **35**(12), 61–70 (1992)

8. Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40**(3), 56–58 (1997)

9. Pazzani, M.J.: A framework for collaborative, content-based and demographic filtering. Artif. Intell. Rev. **13**(5–6), 393–408 (1999)

10. Burke, R.: Hybrid recommender systems: survey and experiments. User Model. User-Adap. Inter. **12**(4), 331–370 (2002)
11. Sill, J., Takács, G., Mackey, L., Lin, D.: Feature-weighted linear stacking. arXiv preprint arXiv:0911.0460 (2009)
12. Tinghuai, M., Jinjuan, Z., Meili, T., Yuan, T., Abdullah, A.D., Mznah, A.R., Sungyoung, L.: Social network and tag sources based augmenting collaborative recommender system. IEICE Trans. Inf. Syst. **98**(4), 902–910 (2015)
13. O'Donovan, J., Smyth, B.: Trust in recommender systems. In: Proceedings of the 10th International Conference on Intelligent User Interfaces, pp. 167–174. ACM (2005)
14. Gao, P., Miao, H., Baras, J.S., Golbeck, J.: Star: semiring trust inference for trust-aware social recommenders. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 301–308. ACM (2016)
15. Bansal, T., Belanger, D., McCallum, A.: Ask the gru: Multi-task learning for deep text recommendations. arXiv preprint arXiv:1609.02116 (2016)
16. Covington, P., Adams, J., Sargin, E.: Deep neural networks for youtube recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 191–198. ACM (2016)
17. Koren, Y., Bell, R., Volinsky, C., et al.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
18. Ge, H., Caverlee, J., Lu, H.: Taper: a contextual tensor-based approach for personalized expert recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 261–268. ACM (2016)
19. Kim, D., Park, C., Oh, J., Lee, S., Yu, H.: Convolutional matrix factorization for document context-aware recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 233–240. ACM (2016)
20. Karatzoglou, A., Amatriain, X., Baltrunas, L., Oliver, N.: Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 79–86. ACM (2010)
21. Koren, Y.: The bellkor solution to the netflix grand prize. Netflix Prize Documentation **81**, 1–10 (2009)
22. Powers, D.M.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation (2011)
23. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 257–260. ACM (2010)
24. Swearingen, K., Sinha, R.: Beyond algorithms: An hci perspective on recommender systems. In: ACM SIGIR 2001 Workshop on Recommender Systems, vol. 13, Citeseer, pp. 1–11 (2001)
25. Knijnenburg, B.P., Reijmer, N.J., Willemsen, M.C.: Each to his own: How different users call for different interaction methods in recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys 2011, pp. 141–148. ACM, New York (2011)
26. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user?s perspective: survey of the state of the art. User Model. User-Adap. Inter. **22**(4–5), 317–355 (2012)
27. He, C., Parra, D., Verbert, K.: Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. Expert Syst. Appl. **56**, 9–27 (2016)

28. Calero Valdez, A., Özdemir, D., Yazdi, M.A., Schaar, A.K., Ziefle, M.: Orchestrating collaboration-using visual collaboration suggestion for steering of research clusters. Procedia Manufact. **3**, 363–370 (2015)
29. Calero Valdez, A., Ziefle, M., Verbert, K.: HCI for recommender systems: the past, the present and the future. In: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 2016, pp. 123–126. ACM, New York (2016)
30. Wiesner, M., Pfeifer, D.: Health recommender systems: concepts, requirements, technical basics and challenges. Int. J. Environ. Res. Public Health **11**(3), 2580–2607 (2014)
31. Farrell, R.G., Danis, C.M., Ramakrishnan, S., Kellogg, W.A.: Intrapersonal retrospective recommendation: lifestyle change recommendations using stable patterns of personal behavior. In: Proceedings of the First International Workshop on Recommendation Technologies for Lifestyle Change (LIFESTYLE 2012), p. 24. Ireland, Citeseer, Dublin (2012)
32. Roitman, H., Messika, Y., Tsimerman, Y., Maman, Y.: Increasing patient safety using explanation-driven personalized content recommendation. In: Proceedings of the 1st ACM International Health Informatics Symposium, pp. 430–434. ACM (2010)
33. Rokicki, M., Herder, E., Demidova, E.: Whats on my plate: towards recommending recipe variations for diabetes patients. In: Proceedings of UMAP 2015 (2015)
34. Elsweiler, D., Harvey, M., Ludwig, B., Said, A.: Bringing the healthy into food recommenders. In: DRMS Workshop (2015)
35. Fernandez-Luque, L., Karlsen, R., Vognild, L.K.: Challenges and opportunities of using recommender systems for personalized health education. In: MIE, pp. 903–907 (2009)
36. Sezgin, E., Özkan, S.: A systematic literature review on health recommender systems. In: E-Health and Bioengineering Conference (EHB), pp. 1–4. IEEE (2013)
37. Achananuparp, P., Weber, I.: Extracting food substitutes from food diary via distributional similarity. arXiv preprint arXiv:1607.08807 (2016)
38. Freyne, J., Berkovsky, S.: Recommending food: reasoning on recipes and ingredients. In: Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 381–386. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13470-8_36
39. Ge, M., Ricci, F., Massimo, D.: Health-aware food recommender system. In: Proceedings of the 9th ACM Conference on Recommender Systems, pp. 333–334. ACM (2015)
40. Harvey, M., Ludwig, B., Elsweiler, D.: Learning user tastes: a first step to generating healthy meal plans. In: First International Workshop on Recommendation Technologies for Lifestyle Change (lifestyle 2012), Citeseer, p. 18 (2012)
41. Berkovsky, S., Freyne, J.: Group-based recipe recommendations: analysis of data aggregation strategies. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 111–118. ACM (2010)
42. Said, A., Tikk, D., Shi, Y., Larson, M., Stumpf, K., Cremonesi, P.: Recommender systems evaluation: A 3d benchmark. In: ACM RecSys 2012 Workshop on Recommendation Utility Evaluation: Beyond RMSE, Dublin, Ireland, pp. 21–23 (2012)
43. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)
44. Ziefle, M., Rocker, C., Holzinger, A.: Medical technology in smart homes: exploring the user's perspective on privacy, intimacy and trust. In: 2011 IEEE 35th Annual Computer Software and Applications Conference Workshops (COMPSACW), pp. 410–415. IEEE (2011)

45. Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R., Zhang, Y.C.: Solving the apparent diversity-accuracy dilemma of recommender systems. Proc. Natl. Acad. Sci. **107**(10), 4511–4515 (2010)
46. Beale, R.: Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. Int. J. Hum. Comput. Stud. **65**(5), 421–433 (2007)
47. Hors-Fraile, S., Benjumea, F.J.N., Hernández, L.C., Ruiz, F.O., Fernandez-Luque, L.: Design of two combined health recommender systems for tailoring messages in a smoking cessation app. arXiv preprint arXiv:1608.07192 (2016)
48. Ekstrand, J.D., Ekstrand, M.D.: First do no harm: Considering and minimizing harm in recommender systems designed for engendering health. In: Engendering Health Workshop at the RecSys 2016 Conference (2016)
49. Mobasher, B., Burke, R., Bhaumik, R., Williams, C.: Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness. ACM Trans. Internet Technol. (TOIT) **7**(4), 23 (2007)
50. Grasgruber, P., Sebera, M., Hrazdira, E., Hrebickova, S., Cacek, J.: Food consumption and the actual statistics of cardiovascular diseases: an epidemiological comparison of 42 European countries. Food Nutr. Res. **60** (2016). doi:10.3402/fnr.v60.31694
51. Spinuzzi, C.: The methodology of participatory design. Tech. Commun. **52**(2), 163–174 (2005)
52. Ekstrand, M.D., Willemsen, M.C.: Behaviorism is not enough: better recommendations through listening to users. In: Proceedings of the 10th ACM Conference on Recommender Systems. RecSys 2016, pp. 221–224. ACM, New York (2016)
53. Barnes, S.B.: A privacy paradox: Social networking in the united states. First Monday **11**(9) (2006)
54. Martin, R.L.: The design of business: why design thinking is the next competitive advantage. Harvard Business Press (2009)
55. Dwork, C.: Differential privacy: a survey of results. In: Agrawal, M., Du, D., Duan, Z., Li, A. (eds.) TAMC 2008. LNCS, vol. 4978, pp. 1–19. Springer, Heidelberg (2008). doi:10.1007/978-3-540-79228-4_1
56. Aggarwal, C.C., Philip, S.Y.: A general survey of privacy-preserving data mining models and algorithms. In: Aggarwal, C.C., Yu, P.S. (eds.) Privacy-Preserving Data Mining, pp. 11–52. Springer, US (2008)
57. Ziefle, M., Halbey, J., Kowalewski, S.: Users willingness to share data on the internet: Perceived benefits and caveats. In: Proceedings of the International Conference on Internet of Things and Big Data (IoTBD 2016), pp. 255–265 (2016)
58. Kowalewski, S., Ziefle, M., Ziegeldorf, H., Wehrle, K.: Like us on facebook! analyzing user preferences regarding privacy settings in germany. Procedia Manuf. **3**, 815–822 (2015)
59. Akhter, H.S.: Privacy concern and online transactions: the impact of internet self-efficacy and internet involvement. J. Consum. Mark. **31**(2), 118–125 (2014)
60. Fogel, J., Nehmad, E.: Internet social network communities: risk taking, trust, and privacy concerns. Comput. Hum. Behav. **25**(1), 153–160 (2009)
61. Freestone, O., Mitchell, V.: Generation y attitudes towards e-ethics and internet-related misbehaviours. J. Bus. Ethics **54**(2), 121–128 (2004)
62. Nissenbaum, H.: A contextual approach to privacy online. Daedalus **140**(4), 32–48 (2011)
63. Wilkowska, W., Ziefle, M.: Privacy and data security in e-health: requirements from the users perspective. Health Inf. J. **18**(3), 191–201 (2012)

64. Ziefle, M., Himmel, S., Wilkowska, W.: When your living space knows what you do: acceptance of medical home monitoring by different technologies. In: Holzinger, A., Simonic, K.-M. (eds.) USAB 2011. LNCS, vol. 7058, pp. 607–624. Springer, Heidelberg (2011). doi:10.1007/978-3-642-25364-5_43

65. Schmidt, T., Philipsen, R., Ziefle, M.: From V2X to Control2Trust. In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2015. LNCS, vol. 9190, pp. 570–581. Springer, Heidelberg (2015). doi:10.1007/978-3-319-20376-8_51

66. Lipkus, I.M., Hollands, J.: The visual communication of risk. J. National Cancer Inst. Monogr. **25**, 149–163 (1998)

67. Seipp, K., Ochoa, X., Gutiérrez, F., Verbert, K.: A research agenda for managing uncertainty in visual analytics. Mensch und Computer 2016-Workshopband (2016)

68. Gigerenzer, G., Edwards, A.: Simple tools for understanding risks: from innumeracy to insight. BMJ. Brit. Med. J. **327**, 741–744 (2003)

69. Sedlmair, M., Meyer, M., Munzner, T.: Design study methodology: reflections from the trenches and the stacks. IEEE Trans. Vis. Comput. Graph. **18**(12), 2431–2440 (2012)

70. Calero Valdez, A., Bruns, S., Greven, C., Schroeder, U., Ziefle, M.: What do my colleagues know? dealing with cognitive complexity in organizations through visualizations. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2015. LNCS, vol. 9192, pp. 449–459. Springer, Heidelberg (2015). doi:10.1007/978-3-319-20609-7_42

71. Parra, D., Brusilovsky, P., Trattner, C.: See what you want to see: visual user-driven approach for hybrid recommendation. In: Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 235–240. ACM (2014)

72. Calero Valdez, A., Brauner, P., Ziefle, M., Kuhlen, T.W., Sedlmair, M.: Human factors in information visualization and decision support systems. Mensch und Computer 2016-Workshopband (2016)

73. Holzinger, A.: Interactive machine learning (iml). Informatik-Spektrum **39**(1), 64–68 (2016)

74. Kieseberg, P., Malle, B., Frühwirt, P., et al.: Brain Inf. **3**, 269 (2016). doi:10.1007/s40708-016-0046-2

75. Malle, B., Kieseberg, P., Weippl, E., Holzinger, A.: The right to be forgotten: towards machine learning on perturbed knowledge bases. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-ARES 2016. LNCS, vol. 9817, pp. 251–266. Springer, Heidelberg (2016). doi:10.1007/978-3-319-45507-5_17

76. Kieseberg, P., Weippl, E., Holzinger, A.: Trust for the doctor-in-the-loop. European Research Consortium for Informatics and Mathematics (ERCIM) News: Tackling Big Data in the Life Sciences 104(1), 32–33

77. Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., Reiterer, S.: Toward the next generation of recommender systems: applications and research challenges. In: Tsihrintzis, G.A. (ed.) Multimedia Services in Intelligent Environments, pp. 81–98. Springer, Switzerland (2013)

78. Rossetti, M., Stella, F., Zanker, M.: Contrasting offline and online results when evaluating recommendation algorithms. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 31–34. ACM (2016)

79. Karumur, R.P., Nguyen, T.T., Konstan, J.A.: Exploring the value of personality in predicting rating behaviors: a study of category preferences on movielens. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 139–142. ACM (2016)

80. Azaria, A., Hong, J.: Recommender systems with personality. In: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 2016, pp. 207–210. ACM, New York (2016)
81. Shiv, B., Carmon, Z., Ariely, D.: Placebo effects of marketing actions: consumers may get what they pay for. J. Mark. Res. **42**(4), 383–393 (2005)
82. Laufer, P., Wagner, C., Flöck, F., Strohmaier, M.: Mining cross-cultural relations from wikipedia: a study of 31 european food cultures. In: Proceedings of the ACM Web Science Conference, p. 3. ACM (2015)
83. Babitsch, B., Braun, T., Borde, T., David, M.: Doctor's perception of doctor-patient relationships in emergency departments: what roles do gender and ethnicity play? BMC Health Serv. Res. **8**(1), 1 (2008)
84. Knijnenburg, B.P., Sivakumar, S., Wilkinson, D.: Recommender systems for self-actualization. In: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 2016, pp. 11–14. ACM, New York (2016)