NUTRITION ASSISTANT APPLICATION

TEAMID:PNT2022TMID34668

TEAM LEADER:SAHAYA NISHIMA D

TEAM MEMBER:SOBHRA RS

TEAM MEMBER:RENI SHAJILA S

TEAM MEMBER:SWEETLIN SAJONI JOY J

Effects and challenges of using a nutrition assistance system:

Healthy nutrition contributes to preventing non-communicable and diet-related diseases. Recommender systems, as an integral part of mHealth technologies, address this task by supporting users with healthy food recommendations. However, knowledge about the effects of the long-term provision of health-aware recommendations in real-life situations is limited. This study investigates the impact of a mobile, personalized recommender system named Nutrilize. Our system offers automated personalized visual feedback and recommendations based on individual dietary behaviour, phenotype, and preferences. By using quantitative and qualitative measures of 34 participants during a study of 2–3 months, we provide a deeper understanding of how our nutrition application affects the users' physique, nutrition behaviour, system interactions and system perception. Our results show that Nutrilize positively affects nutritional behaviour (conditional $R_2=.342$R2=.342) measured by the optimal intake of each nutrient. The analysis of different application features shows that reflective visual feedback has a more substantial impact on healthy behaviour than the recommender (conditional $R_2=.354$R2=.354). We further identify system limitations influencing this result, such as a lack of diversity, mistrust in healthiness and personalization, real-life contexts, and personal user characteristics with a qualitative analysis of semi-structured in-depth interviews. Finally, we discuss general knowledge acquired on the design of personalized mobile nutrition recommendations by identifying important factors, such as the users' acceptance of the recommender's taste, health, and personalization.

Motivation and objective

The research in recommender systems has been recently interested in food recommender systems addressing, among others, nutritional health with different approaches (Trattner and Elsweiler 2018). These systems have the potential to help users navigate the growing amount of multimedia food content (Min et al. 2019a) while fostering healthy eating patterns. Conventional recommender systems learn the users' preferences and try to cater to them, which might enforce recommendations for unhealthy food as well (Elsweiler and Harvey 2015; Schäfer et al. 2017b). Thus, health-aware recommender systems need to also incorporate different parameters related to taste and health into their systems (Elsweiler et al. 2015a, 2016; Ge et al. 2015a). The use of nutrition assistance systems is promising since previous studies have shown that persuasive technologies can help people to eat healthier (Orji and Moffatt 2016).

Existing national and international dietary guidelines constitute important informational sources for nutrition (Painter et al. 2002) but are often based on guidelines for a whole population instead of dealing with individual requirements. Yet, efforts towards personalised nutrition recommendations have been made by Zeevi et al. (2015) and within the Food4Me study (Celis-Morales et al. 2015). Zeevi et al. (2015) incorporated different individual aspects into their algorithm based on machine-learning techniques like dietary behaviour, anthropometrics, blood biomarkers and the gut microbiome. Based on this, they could successfully predict the post-prandial glucose response (PPGR) that varies greatly between different individuals for the same meal. Celis-Morales et al. (2015) examined in their 6-month study the effectiveness of personalised nutrition advice, which was based on dietary, phenotypic and genotypic information. Their results showed higher effectiveness in changing nutrition habits through personalised dietary advice than conventional dietary advice.

Evaluations of recommender systems often focused on measuring algorithmic accuracy, which is insufficient in explaining user experience (Knijnenburg et al. 2012). Further, studies on recommendations and mobile applications frequently cover short-term usage, i.e. think-aloud lab studies of multiple hours or surveys comparing recommender algorithms on a quantitative level. Especially in the context of health recommender systems, it is important to not only evaluate recommendation accuracy. Since healthy recommendations might contrast user preferences, it is crucial to evaluate user satisfaction and changes in behaviour and health over a longer period of time, i.e. multiple weeks, (Schäfer et al. 2017b). Traditional nutrition interventions require months to show lasting effects on nutritional behaviour and physique. Therefore, we evaluate our proposed system, *Nutrilize*, based on a 2–3 months study using a mixed-methods evaluation of the system effects and the user experience. **It is our goal to show how long-term usage of a nutrition assistance system influences the users' (a) physique, (b) nutrition behaviour, (c) system interaction and (d) system perception.** Furthermore, we want to gain insights into the reasons for observations appearing in long-term but not in short-term usage by analysing semi-structured in-depth interviews.

Background

This section discusses related work in four different areas relevant to the application and study design of the presented work: Personalized Nutrition, Food Recommender Systems, Behaviour Change Interfaces, and Evaluation of Recommender Systems. Additionally, we present how our prior work on this system fulfils different prerequisites of the layered evaluation framework of Paramythis et al. (2010).

Personalized nutrition

Food recommendations for nutrition personalized according to individual health requirements are a major research gap identified by several food recommender systems surveys (Trattner and Elsweiler 2017a; Mauch et al. 2018; Min et al. 2019a, b; Theodoridis et al. 2019). One of the most prominent studies on personalized nutrition is the Food4Me study (Celis-Morales et al. 2015). Over an intervention time of six months, 1607 participants across multiple European study-centres received four types of advice via e-mail: 1) a control group receiving conventional advice, 2) a group receiving personal advice based on dietary intake, 3) a group receiving personal advice based on dietary intake and phenotype, 4) a group receiving personal advice based on dietary intake, phenotype, and genotype. Regarding the impact on dietary behaviour, the study shows higher scores according to the MedDiet Mediterranean diet (MedDiet) (Davis et al. 2015) for the personalized advice group than for the control group (Food4Me Study 2016). The phenotypic and genotypic groups showed no significant difference to the group with personalized feedback based on dietary intake (Food4Me Study 2016). While the personalization was more effective, the overall improvement in MedDiet scores was only modest (Food4Me Study 2016). The authors suggest continuous internet-based delivery of advice to increase the impact (Food4Me Study 2016). Similarly, a second evaluation focused on the improvement of the Healthy-Eating-Index shows that participants receiving personalized feedback consumed less red meat, less salt, less saturated fat, more folate, and had higher Healthy-Eating-Index scores at month six than the control arm (Celis-Morales et al. 2016), with no significant difference between the personalization branches. In line with these insights, the *Nutrilize* system is integrating personalization according to intake history, phenotype, and blood measures.

Food recommender systems

Food recommender systems have been implemented and evaluated using many different algorithms and evaluation methods. Most algorithmic solutions are based on standard content-based or collaborative filtering methods, as shown in the overview by Trattner and Elsweiler (2017a). While work with smaller user samples concludes that content-based methods are of superior performance (Freyne and Berkovsky 2010a; Harvey et al. 2013), experiments on larger samples show a higher performance of collaborative approaches (Trattner and Elsweiler 2019). In general, different food recommender systems tested by Trattner and Elsweiler (2019) show lower performance, i.e. area under the ROC curve, than similar systems addressing movies or e-commerce tested by Rendle et al. (2012). Approaches to health-focused food recommender systems have used energy balancing (Ge et al. 2015b), distance from an estimated nutritional requirements (Elsweiler et al. 2015b), or re-weighting according to health metrics (Trattner and Elsweiler 2017b). Beyond prediction of ratings or ranking based on an existing set of recipes, other efforts have been focused on suggesting healthier ingredient substitutes (Achananuparp and Weber 2016) or on generating healthier pseudo recipes (Chen et al. 2019). The implemented algorithm in the *Nutrilize* system is a content-based approach for both health and taste estimation that integrates both energy balancing and nutritional requirements.

Behaviour change interfaces

Beyond the accurate ranking of personalized healthy and tasty recipes, the design of the user interface is a crucial element determining the acceptance of recommendations and the change in behaviour. According to Chen et al. (2017), it is beneficial from a nutritional care perspective to provide information on nutrients and energy in apps. Current commercial nutrition apps mostly focus on the provision of calorie counting and macronutrient distributions such MyFitnessPal (Google LLC 2019a), Yazio (Google LLC 2019f), MyNetDiary (Google LLC 2019b) or MyPlate (Google LLC 2019e). Only a few use colour- and category-based information to ease the food item choices of users and provide educational content on nutrients such as Fooducate (Google LLC 2019c) or LifeSum (Google LLC 2019d). *Nutrilize* goes beyond common mHealth systems for balanced nutritional advice because of its level of personalization included in the underlying algorithm for generating nutritional advice and its design of feedback. Our concept for developing feedback strategies in *Nutrilize* relies on several practical as well as theoretical considerations. We draw on scientific experience from Front-of-pack nutrition labelling (FoPL) schemes using traffic light schemes that proved to be effective for food item choice support in shopping environments (Malam et al. 2009; Dunford et al. 2014; Koenigstorfer et al. 2014; Julia and Hercberg 2017). From a theoretical perspective, our implementation of (visual) feedback bases on the principles of reflective practice as proposed by Schön (1983) in terms of feedback offering reflection on ongoing food choices as well as past nutrition behaviour to optimally support dietary choices. The offered types of (visual) feedback in *Nutrilize* according to Schön (1983) are reflection-in-action implemented in the home screen through colour-coded advice symbols and the food details screen through simulative/predictive feedback on how selected food intake affects individual critical nutrients. The concept of reflection-on-action is offered by different retrospective views on single individual nutrient levels and overall energy in a daily, weekly, and monthly view. Additionally, the system design is oriented by Persuasive System Design as proposed by Oinas-Kukkonen and Harjumaa (2009). Based on their model of persuasive system design (PSD model), *Nutrilize* implements the following PSD elements in its system components: personalization, self-monitoring, suggestion, and simulation (Terzimehić et al. 2016). We further introduce our considerations for feedback implementation in the context of recommender systems by the taxonomy of Nunes and Jannach (2017). They conducted a review on explanations in recommender and decision support systems and ranked visualization of input parameters after natural language-based texts to the second-most used explanation. Further, they rank different purposes of using explanations in studies according to use frequency. The most frequent purpose is transparency, followed by effectiveness, trust, persuasiveness, satisfaction, and education. By integrating textual and visual explanations into our system *Nutrilize*, we tried to increase healthier dietary decisions (efficiency) as well as to make the parameters for calculating recipe suggestions visible (transparency), which again should result in higher levels of trust towards the automated suggestions of the system. Since *Nutrilize* offers nutrient information on a very fine-grained level, with textual and visual feedback, it provides educational potential towards an improved understanding of foods and relating nutrients.

Evaluation of recommender systems

Food recommender systems are most frequently evaluated in offline comparisons of algorithms on a benchmark dataset (e.g. Trattner and Elsweiler 2019). Such offline evaluations often assess a combination of different evaluation metrics for recommender systems (Vargas and Castells 2011) such as accuracy, diversity, and novelty. For recommendations in the food domain, diversity is of special importance for user satisfaction since food choices are recurrent and frequent. *Nutrilize* generates new recommendations each day for all meals. These recommendations should provide diversity on two different levels: within a recommendation and between recommendations over time. A similar evaluation of diversity within and between recommendations has been conducted by van Schaik (2019), who proposed package recommendations for healthy meal plans. Intra-List Diversity (ILD) (Vargas and Castells 2011) is a fitting metric to measure diversity in the context of meal recommendations. The diversity of recommendations over time is relevant in the food context since food decisions are recurrent. To measure diversity over time, we use the Self-System Diversity (SSD) (Vargas and Castells 2011). The SSD considers two subsequent recommendations, which corresponds to two subsequent days in the case of *Nutrilize* recommendations. We additionally propose the Weighted Self-System Diversity (WSSD). It considers more subsequent recommendations and employs the weighting approach proposed by Ding and Li (2005) to model the gradual forgetting of past experiences, a notion presented by Koychev and Schwab (2000). While offline evaluation is frequent and important, online evaluation such as surveys (Musto et al. 2020) and user studies (Massimo et al. 2017) have become more common in the area of food recommender systems. While many user studies on food recommender systems are conducted in shorter sessions (Massimo et al. 2017), the health-context often requires longer durations to show the behavioural and physical impact. One limitation of studying the behaviour within such a setting is that many users behave differently when they are part of a study (known as the Hawthorne effect (McCarney et al. 2007)). Two similar approaches to *Nutrilize* are provided by Alrige and Chatterjee (2018) and Fallaize et al. (2019). To date, Alrige and Chatterjee (2018) have conducted a quality assessment of the mobile app without measuring the health-impact. The eNutri app from Fallaize et al. (2019) which addresses the level of personalization in nutrition-related apps as proposed by Chen et al. (2018) and Franco et al. (2016) is close to ours in using the validated Food4Me Food Frequency Questionnaire (FFQ) (Food4Me 2016) as a basis for offering automated personalized nutrition advice. However, the eNutri app does not offer an automated personalized recipe or food item recommendations but only a personalized report. The eNutri app (Fallaize et al. 2019) has been evaluated by asking users for input and feedback as well as comparing the output to nutrition expert advice. We evaluated *Nutrilize* in several focus groups, simulations, user tests, and a pilot study to verify its validity before showing its impact on perception and interaction in a short-term study. The work continues this line of evaluation by focussing on behavioural and physical changes in a long-term study.

Prior work

The components of the *Nutrilize* system have been evaluated in several unpublished pre-studies, a published design concept (Terzimehić et al. 2016), a published pilot study (Leipold et al. 2018) and a published short-term study on ability-based personalization (Schäfer and Willemsen 2019). This incremental validation of *Nutrilize* can be presented in terms of the framework of layered evaluation by Paramythis et al. (2010). This framework consists of five layers: (a) Collection of input data (CID), (b) Interpretation of the collected data (ID), (c) Modelling of the current state of the “world" (MW), (d) Deciding upon adaptation (DA), and (e) applying (or instantiating) adaptation (AA). Beyond these five layers, Paramythis et al. (2010) present suggestions for evaluating adaptation as a whole. Regarding the first layer, the *Nutrilize* application collection input data on the food behaviour of users and on the interactions with the different screens. Regarding the quality of food items, we compared three different nutritional databases regarding the accuracy and variety of nutritional information. Regarding the nutritional values of recipes, we tested our matching of recipes to the BLS database with a test set of ground truth samples provided by nutrition scientists. For the tracking of user interactions, we verified the completeness and validity of our tracking system during our pilot study (Leipold et al. 2018). Regarding the second layer of interpretation of the collected data, we simulated both the health recommendations as well as the taste recommendations for prototypical user inputs and let the results be verified by a potential user, in the taste case, and nutrition scientists, in the health case. For the third layer of accurately modelling the world, we used focus groups, interviews, and user tests using the think-aloud method to verify the appropriateness and understandability of our system. On the fourth layer of choosing the fitting form of adaptation, we performed to date only one comparative study that tests the effectiveness and perception of two different levels of personalization (Schäfer and Willemsen 2019). Finally, for the fifth layer, we conducted a full pilot study over three weeks to measure the systems robustness as well as the user's interactions and perceptions (Leipold et al. 2018). The long-term study presented in this paper can either be described as an evaluation of the system as a whole instead of a layer-wise or piece-wise evaluation of specific factors. Many of the components of our system are interconnected and cannot be measured holistically in separate evaluations. For example, the quality of real user input influences the predictive quality, which in turn changes the impact of the persuasive and personalized feedback that the system provides. Thus, during our long-term study we want to collect a variety of in-detail assessments of a real-life usage scenario.

The *Nutrilize* application

We examined our research questions using the nutrition assistance system *Nutrilize*. The early design and background of *Nutrilize* is described by Terzimehić et al. (2016) and a pilot study on the general usage of *Nutrilize* is reported in Leipold et al. (2018). Schäfer and Willemsen (2019) show the most current version of the system, but with a Rasch-based tailoring and a Dutch target group. This section will shortly describe all the features of the current system version used during our long-term study. First, we describe all features required for tracking the daily dietary intake of the participants, namely the food-search, food-details, sports-search, and diary. Second, we describe the recommendation features. Third, we describe all visual feedback screens, namely the statistics screen, nutrition status screen, home screen, and energy overview. Finally, we show all the administrative features such as the preference screen, the profile screen, the login screen, and the settings screen.

Tracking features

The diet tracking of each user is done using a search interface). The user can either perform a free text search, select the food item from a tree structure, or select one of his/her recent or favourite items. When choosing a food item, the user can either directly add the default amount by clicking on the plus button, or first, click on the food item to receive some more detailed information on the food's nutrients and choose a custom portion size to add to the diary). The same mechanism is given for entering physical activity). Instead of portion sizes, here the user should choose the amount of time for the performed physical activity. Finally, the user can review and update all his entered food items in the diary. Here he can also enter food for past days or delete wrong entries .