Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching

Buzcu, Berk; Varadhajaran, Vanitha; Tchappi, Igor; Najjar, Amro; Calvaresi, Davide; Aydoğan, Reyhan

DOI
10.1007/978-3-031-21203-1_2

Publication date
2023

Document Version
Final published version

Published in
PRIMA 2022

Citation (APA)
Buzcu, B., Varadhajaran, V., Tchappi, I., Najjar, A., Calvaresi, D., & Aydoğan, R. (2023). Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching. In R. Aydoğan, N. Criado, V. Sanchez-Anguix, J. Lang, & M. Serramia (Eds.), *PRIMA 2022: Principles and Practice of Multi-Agent Systems - 24th International Conference, Proceedings* (pp. 20-36). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 13753 LNAI). Springer. https://doi.org/10.1007/978-3-031-21203-1_2

Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching

Berk Buzcu[1(✉)], Vanitha Varadhajaran[2], Igor Tchappi[2], Amro Najjar[3], Davide Calvaresi[4], and Reyhan Aydoğan[1,5]

[1] Computer Science, Özyeğin University, Istanbul, Turkey
berk.buzcu@ozu.edu.tr , reyhan.aydogan@ozyegin.edu.tr
[2] University of Luxembourg, Esch-sur-Alzette, Luxembourg
igor.tchappi@uni.lu
[3] Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxemburg
amro.najjar@list.lu
[4] University of Applied Sciences and Arts Western Switzerland, Delémont, Switzerland
davide.calvaresi@hevs.ch
[5] Interactive Intelligence, Delft University of Technology, Delft, The Netherlands
https://expectation.ehealth.hevs.ch/

Abstract. People's awareness about the importance of healthy lifestyles is rising. This opens new possibilities for personalized intelligent health and coaching applications. In particular, there is a need for more than simple recommendations and mechanistic interactions. Recent studies have identified nutrition virtual coaching systems (NVC) as a technological solution, possibly bridging technologies such as recommender, informative, persuasive, and argumentation systems. Enabling NVC to explain recommendations and discuss (argument) dietary solutions and alternative items or behaviors is crucial to improve the transparency of these applications and enhance user acceptability and retain their engagement. This study primarily focuses on virtual agents personalizing the generation of food recipes recommendation according to users' allergies, eating habits, lifestyles, nutritional values, etc. Although the agent would nudge the user to consume healthier food, users may tend to object in favor of tastier food. To resolve this divergence, we propose a user-agent negotiation interacting over the revision of the recommendation (via feedback and explanations) or convincing (via explainable arguments) the user of its benefits and importance. Finally, the paper presents our initial findings on the acceptability and usability of such a system obtained via tests with real users. Our preliminary experimental results show that the majority of the participants appreciate the ability to express their feedback as well as receive explanations of the recommendations, while there is still room for improvement in the persuasiveness of the explanations.

Supported by Chist-Era (grant CHIST-ERA-19-XAI-005).

R. Aydoğan et al. (Eds.): PRIMA 2022, LNAI 13753, pp. 20–36, 2023.
https://doi.org/10.1007/978-3-031-21203-1_2

1 Introduction

According to the World Health Organization (WHO), non-communicable diseases (e.g., cardiovascular diseases, chronic respiratory diseases, and diabetes) are responsible for 63% of all deaths worldwide[1]. Moreover, WHO outlines that these diseases are preventable by effectively tackling shared risk factors such as unhealthy diets. Personal preferences and constraints (e.g., cultural, religious, and sustainable diets [7]) and unhealthy derivatives possibly hidden (or overlooked) in a large variety of food items highlight the need for guidance. To this end, food recommender systems are increasingly proposed to guide people in selecting suitable recipes [30]. Indeed, food recommenders have dramatically grown, primarily fueled by globalization (more availability and broader variety) and the rise of ultra-processed food that has skyrocketed metabolic and overweight issues [12].

One may argue that they can find countless recipes over the internet. However, picking the "best" one for a given individual in a specific situation is highly complex. Objectively handling such a vast collection of possibilities and crossing variables such as allergens, nutrition values, personal needs, calories already intake, historical values, and the *preference* of the moment is challenging. Therefore, a personalized support system is needed. Nutrition virtual coaches (NVC) are intended to recommend the most fitting recipes to the user according to a broad set of variables. NVC can, indeed, consider users' health [29] as well as their needs, requests, and historical consistency over time. NVC can support a wide range of goals, ranging from gaining muscles, to weight loss, even for individuals with nutrition-related diseases (i.e., obesity[2]). Such support is envisioned to be educative. Instructing the user would allow reducing the dependency on the NVCs progressively.

Existing solutions both from research [5] and industry [25] tried to cope with such goals. However, they lack clarity and transparency, generating a lack of trust and efficacy. Explainable AI (XAI) techniques have been adopted in several tangential applications domains, such as transportation [20], fleet management [16], neurosciences [9] etc. to bring such transparency. Moreover, some studies have approached explainable food recommendations by proposing a semantic model [21] and incorporating negotiation to gently navigate the user towards a certain quality of life goal [18]. While these works have been considered recommender systems, to the best of our knowledge, no existing system qualifies as a whole Nutrition Virtual Coach allowing the agents to explain the recommendations to the user, and contend interactively over it for the sake of achieving the desired behavior change.

This work presents an interactive and explainable protocol enabling an NVC to promote healthy(-ier) food. To do so, we have developed a simple health score calculation module, a module with a multi-criteria additive utility function to

[1] https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

[2] https://www.cdc.gov/chronicdisease/resources/publications/factsheets/nutrition.html.

rank the recipes logically, and an OWL ontology to classify/relate users and ingredients (best fitting). Moreover, we tested and assessed the protocol with individuals characterized by various backgrounds.

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 presents the explainable argumentation negotiation module for NVC. Section 4 evaluates and discusses the obtained results. Finally, Sect. 5 concludes the paper and outlines future works.

2 Related Works

This section briefly overviews the literature on recommender systems in the context of food recommender systems (Sect. 2.1) and evolving towards explainable and interactive recommendations (Sect. 2.2).

2.1 Conventional Food Recommendation

One of the first concepts of a food recommender system, CHEF, dates back to 1986 [14]. It tests case-based planning, defining a few success/failure conditions and attempting to replace/improve food items within recipes. The cased-based planning model used in CHEF requires an extensive initial knowledge base, remarkable pre-processing, and the creation of plans and backup plans for each recipe. Freyne and Berkovsky implement the general intuition of recommender algorithms such as collaborative filtering (CF), and content-based (CB) approaches to recommend recipes [10]. Their strategy to determine ingredient weights and use them with the CF and CB performs better when making predictions than directly using the recipes. Ge, Ricci and Massimo introduce concepts of personalization of the recommendations subordinating taste to health [11]. Chi, Chen and Tasi focus on recommending food for chronic individuals (i.e., kidney diseases) [6]. The authors architected a specific Ontology Web Language (OWL) ontology embedding health-relevant aspects rather than specific calculations. A more generalized healthy recommendation framework is proposed by Chen *et al.* [5]. It focuses on modifying unhealthy recipes from a dataset of unhealthy food. They propose a new deep learning-based method (IP-embedding) to match recipes with desired ingredients. The IP-Embedding is used to build a pseudo recipe (a set of ingredients forming the desired outcome) from the requirements, which is then matched to healthy ingredients, and finally matched with a real recipe via the MSE metric. Similarly, Teng, Lin and Adamic seeks to build a pointwise comparison metric to understand how to realize recipes from ingredients and swap them with healthier alternatives [27]. Ingredients/food substitution has also been tackled by Elsweiler, Trattner and Harvey [8]. They metricize the nutritional values for fat, sodium, etc., for a predetermined healthy range. Then, images of the recipes push users to prefer healthier food rather than unhealthy options.

2.2 Early Interactive Recommendation System

Explainable AI (XAI) is pervading humans' daily lives. AI predictors and classifiers are no longer allowed to be opaque. To be trusted and make a real impact on humans, they need to be more transparent, understandable, and inspectable [2]. Indeed, recommender systems are expected to equip their outcomes with explanations [13]. Such explanations should allow for justification, control, and discovery of new aspects of the proposed outcome [1]. Padhiar *et al.* propose a food recommender system drafting explanations from knowledge-based ontology [21]. Samih, Adadi and Berrada push the concept further, proposing a knowledge-based explainable recommender system where they generate explanations using the probabilistic soft-logic framework [24]. Finally, to increase the interaction between the user and the virtual assistants, Lawo *et al.* define a cluster of consumers with ethical and social priorities and include them in the recommendation process negotiation as central concerns — recording positive feedback w.r.t. concerning the alternatives proposed [18].

Overall, recommendation systems are not new in the nutrition domain. Pursuing healthy, sustainable, and taste-based combinations are the most targeted goals. Recent studies tried to embed explanations in the recommendations to foster transparency — henceforth trust and acceptability. Aligned with this trend, the following section presents an interactive communication/negotiation protocol enabling explainable mechanisms to dive into specific topics and to learn from the user's feedback promptly.

3 Ontology-Based Negotiation for NVC

The well-known tendency to prefer unhealthy food is worsened by the plethora of internet-sourced popular recipes which foster unhealthy habits [26]. Accordingly, the NVCs aim to improve dietary choices and reduce health risks through personalized recommendations. To do so, there is a need to resolve the conflict between user preferences regarding the food's taste and the recipes' healthiness. Furthermore, the system should explain well-grounded reasons why the given recipe is recommended to the user so that the NVC can persuade its users to improve their eating habits.

This section proposes an NVC leveraging ontology-based reasoning to interact with the user in a turn-based fashion. The goal is to understand their needs and interests, and make healthy recommendations, while still considering their preferences. The interaction between the user and the NVC is governed by the explanation-based negotiation protocol presented in this work. Section 3.1 describes the food ontology required for ontological reasoning. Section 3.2 describes the explanation-based negotiation protocol, Sect. 3.3 explains our basic recommendation algorithm and finally, Sect. 3.4 elaborates on the explanation generation algorithm.

3.1 Food Ontology and Recipe Repository

An OWL-based ontology database is utilized in Protege to represent the relationship between the users and their eating habits and to capture the structural similarities among the ingredients of the recipes. The ontology consists of two main concepts: *User* and *Food*. The *User* concept captures the users eating habits, such as religious or lifestyle restrictions. We define object properties such as *doesNotEat* to determine what kind of food ingredients the user would not consume. Figure 1 shows a small part of the ontology indicating user eating habits. Here, the restrictions (e.g., Muslims do not eat pork) are encoded by connecting the object properties (shown in diamonds) to both the *User* and *Food* ends.

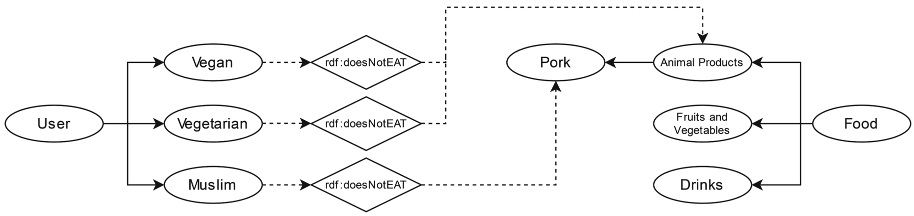


Fig. 1. A sample of the general structure of the OWL based ontology

The *Food* concept involves a hierarchy of food recipe ingredients (e.g., *Beef* is a sub-concept of *Animal Products*, a *Cucumber* is a sub-concept of *Vegetables* etc.). Figure 2 shows the Protege view of some food classifications and some of their ingredients or instances. The agent can query the OWL ontology and decide whether the user can consume the given recipe based on their eating habits.

We realized a repository by incorporating two different datasets: *foodRecSys*[3] and *FoodBase Corpus* [22]. *foodRecSys* contains around 46K recipes with comprehensive nutritional values. It also contains each recipe's name, photo, ingredients, and cooking steps. However, the dataset lacks the structural information our system needs to construct our ingredient concept hierarchy on the ontology. Therefore, we use *FoodBase Corpus* [22], which contains a structured annotation of recipe ingredients from the same source. Using this annotation, we automatically construct the ontology structure. Then, we manually adapted it to fit our recipe repository.

Since the content of the recipes involves more complex structures (e.g., *slice of a tomato* instead of *tomato*), we need to pre-process the ingredients to match our ontological instances. In this process, we apply the Levenshtein Distance[4]

[3] https://www.kaggle.com/datasets/elisaxxygao/foodrecsysv1?resource=download&select=core-data_recipe.csv.

[4] https://blog.paperspace.com/measuring-text-similarity-using-levenshtein-distance/.

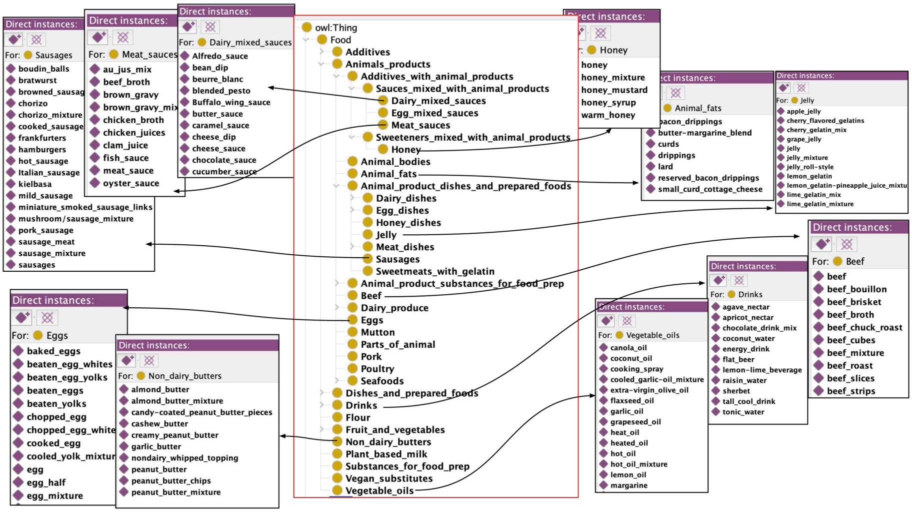


Fig. 2. Protege view of the Food class and some of their ingredients using OWL

to detect the best match. This match is mapped onto our repository, enabling the reasoner to use the food ontology. Furthermore, the recipe repository does not include cuisine information. Thus, we used an additional cuisine dataset[5] to incorporate the cuisine information into the recipe repository. Finally, we filtered the recipes that we could not find the corresponding cuisine information. Ultimately, the remaining number of recipes is 15K.

3.2 Explanation-Based Negotiation Protocol

The bilateral negotiation protocol enables the user to specify their constraints w.r.t. preferences and proactively give feedback while allowing the NVC agent to generate personalized recommendations along with explanations. Note that the explanation is about why the agent chooses a given recommendation. Similar to the Alternating Offer Protocol [3,23], the interaction follows turn-taking fashion (see Fig. 3).

According to the proposed protocol, the user initiates the interaction by sending their constraints (C), which may consist of the ingredients the user may be allergic (e.g., milk, peanuts) to; the (dis)liked ingredients (e.g., specific meat/vegetables); and the desired type of cuisine (e.g., Middle Eastern, Italian, French). After receiving the user's constraints, the agent recommends a recipe (R) along with its explanation (ϵ). Then the user can: Accept R, leave without an agreement, criticize R, ϵ, or both. When the user makes a critique, the agent can revise its recommendation/explanation, regenerating (R'), (ϵ'), or both. This interaction continues in a turn-taking fashion until reaching a termination condition (i.e., Accept or Leave w/o Recommendation) or the time deadline is reached.

[5] https://cosylab.iiitd.edu.in/culinarydb/.

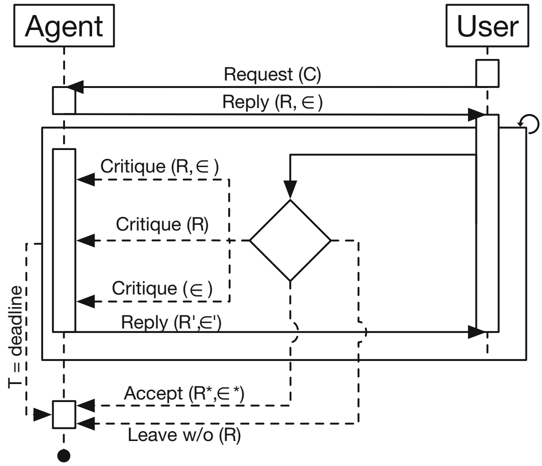


Fig. 3. FIPA description of the negotiation protocol

In our current implementation, a user can criticize the given recommendation by referring to pre-structured critiques as follows, where Y denotes one of the ingredients chosen by the user. (i) I ate Y recently, (ii) I'm allergic to Y, (iii) I don't like Y, and (iv) I want to give custom feedback.

Similarly, the user can criticize the explanations communicated alongside the recommendations with the pre-defined statements such as (i) The explanation is not convincing, (ii) The explanation does not fit my case, (iii) The explanation is incomplete, (iv) The explanation is not clear enough, and (v) I disagree with the explanation.

It is worth mentioning that the protocol is flexible enough to allow any kind of explanation and critique. For simplicity, this work only covers some basic structured phrases as mentioned above.

3.3 The Baseline Recommendation Strategy

Filtering and Scoring Recipes. To analyze the applicability of the designed protocol, we have developed a basic recommendation strategy relying on filtering and scoring the recipes with respect to the user's constraints and healthiness, as seen in Algorithm 1. The agent first filters the recipes according to the user's eating habits/constraints via ontology reasoning on what kind of ingredients the user would not consume (Lines 1–3). Assuming that the user is vegan, then the agent first filters the recipes containing animal-related products. Then, if the same user specifies that they do not like "zucchini", the agent removes the recipes containing zucchini from the remaining candidate list, R_u. In turn, the utilities of the remaining candidate are calculated by considering both healthiness and their alignment with the user preferences. Then, the recipes are ordered according to the calculated utilities (Lines 4–5)[6]. The recipe with the highest utility is taken

[6] The details of the utility calculation are explained below.

as a candidate recipe, and the system retroactively generates an explanation in line with the recipe's properties (Lines 6–7). This candidate recipe and its corresponding explanation are given to the user.

When the agent receives feedback from the user regarding the recipe, F_r, it filters the candidate recipes according to the updated constraints given by the feedback and selects the highest-ranked recipe similarly (Lines 10–15). When the agent receives feedback from the user regarding the explanation, F_ϵ, it simply generates a new explanation with the underlying recipe (Lines 16–18).

Algorithm 1: AgentDecisionFunction

Data: R: Recipes, U: User;
$R_u \subset R$: Recipe dataset tailored for the user;
H_u: Eating habits of the user, P_u: User Constraints/Preferences;
r_c: Candidate recipe, ϵ: Explanation for candidate recipe;
F_r: Feedback to the recipes, F_ϵ: Feedback to the explanation;
1 **if** *firstRecommendation* **then**
2     $R_u \leftarrow$ filterRecipesByCondition(R, H_u);
3     $R_u \leftarrow$ filterRecipesByCondition(R_u, P_u);
4     $U_{R_u} \leftarrow$ calculateUtilities(R_u);
5     $R_u \leftarrow$ rankRecipes(R_u, U_{R_u});
6     $r_c \leftarrow$ getHighestRankRecipe(R_u);
7     $\epsilon \leftarrow$ generateExplanation(r_c);
8 **end**
9 **else**
10   **if** F_r *exists* **then**
11         $R_u \leftarrow$ filterRecipesByCondition(R_u, F_r);
12         $U_{R_u} \leftarrow$ calculateUtilities(R_u);
13         $R_u \leftarrow$ rankRecipes(R_u, U_{R_u});
14         $R_c \leftarrow$ getHighestRankRecipe(R_u);
15   **end**
16   **if** F_ϵ *exists* **then**
17         $\epsilon \leftarrow$ generateExplanation(r_c);
18   **end**
19 **end**
20 return (r_c, ϵ);

Utility Estimation. To select the suitable recipe, this paper relies on a multi-criteria decision-making [17]. Multi-criteria decision analysis allows decisions among multiple alternatives evaluated by several conflicting criteria [31]. In this paper, the multi-criteria decision analysis is done by ranking recipes through a multi-criteria function. The multi-criteria function gives each recipe a score in the dataset. One of the main advantages of using a mathematical function is the transparency of the function and its outcomes. This feature is well suited for our proposed NVC due to the explainability of the generated behavior.

Now let us explain how our agent calculates the overall utility of the recipes. Based on the multi-criteria, the agent considers three criteria: Active Metabolic

Rate (AMR) score, nutrition value score, and users' Satisfaction score. The final score of the recipes is the weighted sum of the score provided by each module as presented by Eq. 1 where w_a, w_n, w_u denote the weights of each AMR score, nutrition value score, and users' satisfaction score, respectively. Note that each score is normalized to ensure that the overall score is ranged within [0, 1].

$$recipeScore = w_n * nutrientsScore + w_a * amrScore + w_u * UsersScore \quad (1)$$

The nutrient-based score is calculated according to the nutritional information of the recipes, such as proteins, lipids, carbohydrates, cholesterol, sodium, and saturated fats. These nutrients have respective recommended amounts for a healthy life [28]. In this work, we take into account the nutrition intake limits specified by the WHO organization[7]. Accordingly, the nutrition-based score is calculated as seen in Eq. 2 where each individual nutrition score is calculated according to Eq. 3. We assume that consuming less than each nutrient's minimum amount (min_n) is better than its maximum amount (max_n). By following this heuristic, the individual score of each nutrient is calculated.

$$\begin{aligned} nutrientScore(recipe) = score(pro) + score(lip) + score(cb) + \\ score(ch) + score(sod) + score(sat) \end{aligned} \quad (2)$$

$$score(n) = \begin{cases} 5 & \text{if } n \in [min_n, max_n] \\ 3 & \text{if } n < min_n \\ 1 & \text{else} \end{cases} \quad (3)$$

AMR is the number of calories that a person must consume daily depending on his height, sex, age, weight, and activity level. Such preliminary information is taken during the registration of the users. The value of AMR is based on the value of Basal Metabolic Rate (BMR), the number of calories required to keep a body functioning at rest, the activity level of the person, and the desire of the person to maintain or reduce his current weight. Table 1 presents the values to keep the current weight. To compute the AMR score based on the minimum and maximum amount of calories required for a given user available in literature [28], we rely on the same assumption of Eq. 3 that is consuming fewer calories than required ($score = 3$) is better than consuming more calories than required ($score = 1$). In addition, when the amount of calories computed is between the minimum and maximum amount of calories, the score is set to 5. Historically, the most used formula to compute BMR is the [15] equation with Eqs. 4 and 5, for men and women respectively. The authors estimated the constants of Eqs. 4 and 5 by several statistical experiments [15].

$$BMR = 10 * weight + 6.25 * height - 5 * age + 5 \quad (4)$$

$$BMR = 10 * weight + 6.25 * height - 5 * age + 161 \quad (5)$$

[7] https://www.who.int/news-room/fact-sheets/detail/healthy-diet, http://www.mydailyintake.net/daily-intake-levels/.

Table 1. Daily recommended kilocalories (kcal) intake to maintain weight [28]

Activity level	Daily calories
Too little exercise	$calories = BMR * 1.2$
Light exercise	$calories = BMR * 1.375$
Moderate exercise	$calories = BMR * 1.55$
Strong exercise	$calories = BMR * 1.725$
Very strong exercise	$calories = BMR * 1.9$

User Satisfaction Score. Lastly, the user satisfaction score is calculated by considering the popularity of the recipe among all users and the current user's preferences equally. For the popularity of the recipe, we use the ratings given by the other users between [1, 5]. These values are normalized to [0, 1]. Meanwhile, regarding the user's preferences, we check how many ingredients are considered to be liked by the user. Here, to determine whether an ingredient is liked or not, we can use the explicit feedback from the user as well as rely on user profiling to predict whether the given ingredient is likely to be preferred to be consumed. Here, we use Jaccard Similarity [4] to estimate the individual user satisfaction (the rate of the number of the preferred ingredients over the number of all the ingredients of a given recipe).

Let us assume the user-submitted his preference for ingredients (i_1, i_2, i_3) and we have a recipe such that $R_1 = i_1, i_2, i_5, i_6$. Each ingredient that exists with the liked constraint is considered to be 1 and 0 otherwise. The mean of this operation is 0.5, which is effectively the score of R_1 for this user. For all the recipes, the scores are then max-normalized to place the values between [0, 1], resulting in a relative level of importance for the given recipe. For instance, let's assume that the system knows that the user likes the ingredients i_1, i_2, and i_3 and calculate the score of a recipe consisting of the following ingredients:i_1, i_2, i_5, i_6. The individual user satisfaction would be 2/4 according to Jaccard similarity. If the overall user rating of that recipe is equal to 4 out of 5, then the overall score would be equal to 0.65 ((0.5+0.8)/2).

3.4 Explanation for Recommendation

We present a straightforward explanation generation approach to demonstrate how the NVC agent interacts with its users. Two types of explanations are considered: health-related and preference-related. Six promising health-related explanations are generated from our dataset by taking the nutritional values into consideration, such as protein, vitamin coverage, and cholesterol counts. If the recommended recipe satisfies the given health conditions, we need to identify the main ingredient contributing to this condition, as seen in Table 2, where X is the amount of the nutrition. Note that more sophisticated explanations could be produced by consulting some nutritionists and enhancing ontology reasoning.

Three different explanations are presented regarding the user preferences: chosen cuisine, chosen ingredient, and overall popularity of the recipe (i.e., community score S) as seen in Table 2. To generate such explanations, we match the ingredients of the recipe (i.e., Y, Z) with the corresponding explanation. Finally, the agent generates its complete explanations by combining a fixed starting statement and randomly chosen health and preference-related explanations.

Table 2. Explanation variables and examples

Criteria	Example
Health related explanations	
Protein amount covers user needs for a meal	This recipe contains X grams of protein, which is about X% of your daily requirement. Your body needs proteins from your organs to your muscles and consuming the necessary amount of it is important!
Calorie count is above 30% of BMR	You should eat this food because it covers a good portion of your necessary daily calorie intake by X. You should eat X for a balanced diet and maintain a healthy weight!
Vitamin B amount covers the user needs for a meal	B vitamins have a direct impact on your energy levels, brain function, and cell metabolism, and you should consume X gram for a healthy vitamin B level in a meal!
Vitamin C amount covers the user needs for a meal	Vitamin C is needed for the growth and repair of tissues in all parts of your body. This recipe contains: X% of it and eating it will make you feel more energetic!
The amount of iron must be enough for a meal	Lack of Iron in your system could be critical, causing a disease that is known as iron deficiency anemia. This recipe supplies you with X of it.
The cholesterol must be below a threshold	This recipe has a very low cholesterol count. Cholesterol is linked with a higher risk of cardiovascular disease, and this food is great for low amounts of it.
Preference related explanations	
Users' chosen cuisine matches the recipe cuisine	This recipe is also a part of the cuisine you like: Z
Users' chosen ingredient matches the recipe	This recipe contains the ingredient you wanted: Y_1 and Y_2.
The recipes community score is above a threshold	This recipe was rated S stars by the community, and you might like it too!

4 Evaluation

The acceptability and appreciation of the proposed framework are evaluated via a user study involving 53 participants. This section presents the experimental setup and participants, and discusses the results of the user experiments.

4.1 Experimental Setup

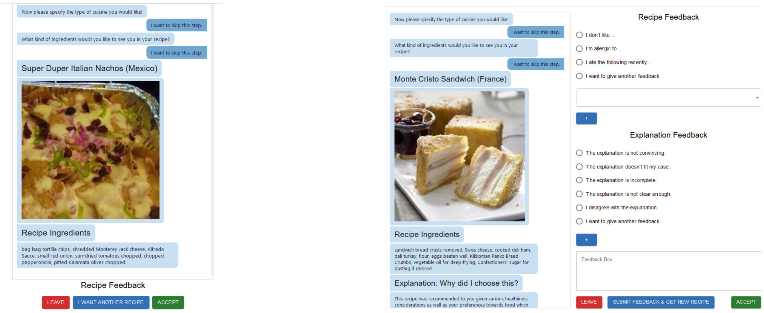
To investigate the effect of interactive explanations and critiques introduced by the proposed protocol, we leverage a Web-based platform allowing the users to experience both the explanation-based negotiation protocol (i.e., the interactive recommender) and its replica without the explanations and critiques component (i.e., a regular recommender) — see Fig. 4.

Prior to the experiments, each user is asked to fill out a pre-survey and registration form to specify their gender, age, height, weight, sports activity level, eating habits, and allergies[8]. This information is used to estimate the healthiness score of recipes recommended to the user (see Sect. 3.3). Seeking to instill different experiences, the two environments (i.e., regular and interactive e-coach/food recommenders) are proposed randomly to each user with a 5-minute break between the two sessions. At the end of the test, the users are asked to fill a questionnaire consisting of mostly 5-point Likert scale questions regarding their experience in both sessions, employing a within-subject design [19].

4.2 Participants

We have recruited 53 users (i.e., 33 men and 20 women with various backgrounds). The mean age is 25.9 with a max of 51 and the min is 19 years old. The participants were asked to rank the importance of five criteria: “Nutritional factors”, “Past experience with taste”, “How it looks”, “Price of the ingredients”, and “Cooking style” while making their decisions on a food recommendation. Figure 5 shows the histogram analysis of this ranking. Twenty-eight users seem to prefer the recipes they knew already. Ten users have chosen nutritional factors among those that they prefer or consider healthier.

[8] We would like to state that the experiment protocol adopted in this study was approved by the Ethics Committee of Özyeğin university, and informed consent was obtained from all participants.



(a) Regular Recommender Session (b) Interactive Recommender Session

Fig. 4. Regular and interactive recommendation sessions

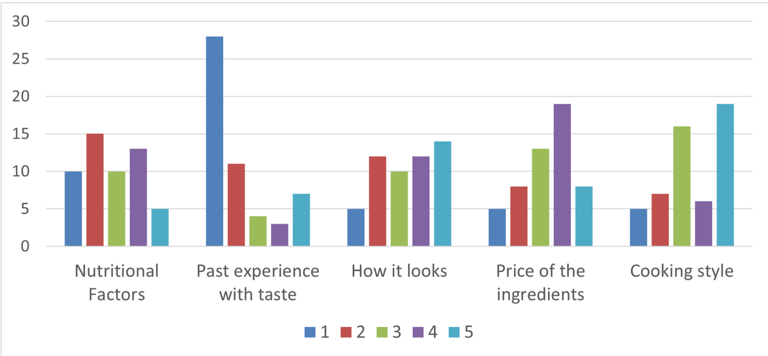
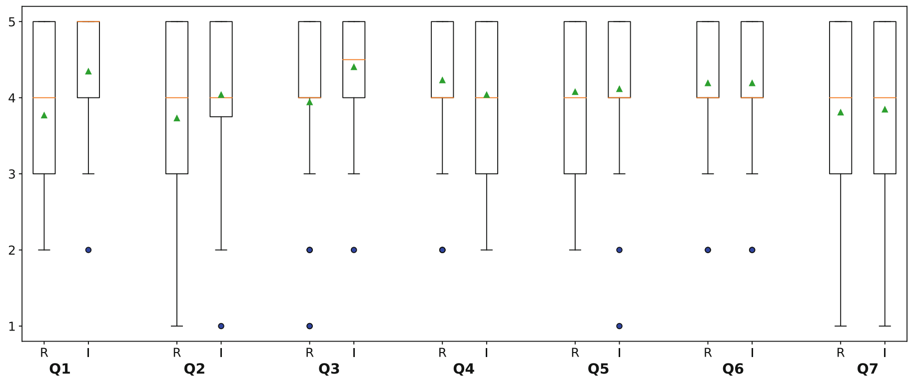


Fig. 5. Pre-survey questionnaire ranking question histogram

4.3 Experimental Results

This section analyzes the findings of the tests and the users' responses to the post-test survey. Since the experiment is composed of two sessions (i.e., Interactive and Regular) and the comparative questions are the same for both, we performed a within-analysis statistical tests. The data is not normally distributed which is one of the main assumptions made by the pairwise T-test. Thus, we apply its corresponding non-parametric test called the Wilcoxon sign rank test [19]. For all tests, the Confidence Interval (CI) is set to 0.95, $\alpha = 1 - CI = 0.05$.

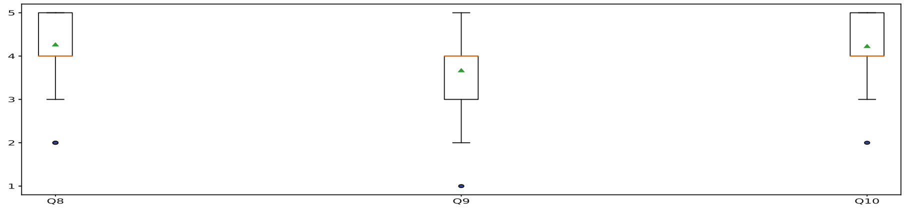
Figure 6 shows the box plot (including the mean) and the p-values of the first set of questions in the experiments between the regular and interactive sessions respectively. The yellow lines represent the median, the triangles in green the means, and the small blue circles the outliers. The results show that for Q1 ($p = 0.001$) and Q2 ($p = 0.011$) the interactive session is significantly better than the regular session. Indeed, these two questions aim at the sociability of the system, and the difference may stem from the engagement provided by the interactive session. Q3 (p = 0.002) is about the completeness of the system. The users answered that the interactive session provided a better set of information than the regular session to make an informed decision. Q4 (p = 0.931), Q5 (p = 0.431), and Q6 (p = 0.506) qualify the usability of the system. Assessing them, we can convene that adding an interactive dimension to the system can still be effective and efficient. Nevertheless, only a minor part of the users (12 out of 53 users or 23% of them) still prefer the regular one over the interactive system.



Id	Questions label	Metric	W	p
Q1	The interaction interface enabled me to effectively express my feedback for the recommendations.	Sociability	52.0	**0.001**
Q2	The recommender system was receptive to my feedback.	Sociability	52.5	**0.011**
Q3	The recommender system gave me enough information to decide.	Completeness	40	**0.002**
Q4	The recommender system was easy to use.	Usability	170.5	0.931
Q5	I found a recipe without consuming too much time.	Usability	157.0	0.436
Q6	The interaction with the system was smooth and effortless.	Usability	138.5	0.506
Q7	I would like to use such a system in my daily life.	Acceptability	34.5	0.356

Fig. 6. Box plot and p-values of the first set of questions

Furthermore, we questioned the users regarding their experiences with the explanations, as illustrated in Fig. 7. Concerning Q8, mean, median, and mode are greater or equal to 4 ("Agree"). Moreover, three quartiles of Q8 are equal or greater than 4 ("Agree") which means most of the users (about 75%) were satisfied with the explanations. Regarding question Q9 related to the usefulness of the explanations, two quartiles are greater or equal to 4 ("Agree") and two quartiles are between 2 ("Disagree") and 3 ("Neutral"). That means at least half of the participants agree with the usefulness of the explanation while others either disagree or are neutral. Finally, concerning Q10, three quartiles are equal or greater than 4 ("Agree"), which means most of the users appreciate receiving explanations in addition to recommendations. Moreover, since the mode is 5, users who strongly agree are greater than those who agree to receive explanations.



Id	Questions label	Metrics	Mode	Median	Mean	std
Q8	The explanations exchanged during the interaction were satisfactory.	Satisfaction	4	4	4.26	0.79
Q9	The explanations for recommendations have impacted my choice.	Usefulness	4	4	3.67	1.02
Q10	Rate your appreciation of the idea of receiving explanation in addition to recommendations.	Usefulness	5	4	4.23	0.80

Fig. 7. Average ratings of the explanation related questionnaire

Finally, the experiment revealed that while users scored the looks of an image to be less important than other factors contributing to their decision, at the end of the experiment, a considerable 55% of them gave the highest score (Strongly Agree) to whether the images influenced their decision or not.

5 Conclusion and Future Work

This study developed an interaction protocol for the XAI-based NVCs to improve the system's transparency via interactive explanations. To this end, it proposes a specialized alternating offers protocol, an OWL-based ontology for ontological reasoning, and a utility function that considers nutritional information to determine the healthiness of a recipe and the user's preferences. Such a contribution has been tested with 53 individuals who, in the final questionnaire, have remarked the concrete contribution conveyed by the interactive explanation-base interaction w.r.t. conventional recommender systems. In future work, we plan to design a more sophisticated explanation mechanism and recommendation strategy to be framed in a reconciling and scalable agent-based framework.

Acknowledgments. This work has been par ally supported by the Chist-Era grant CHIST-ERA-19-XAI-005, and by *(i)* the Swiss National Science Foundation (G.A. 20CH21_195530), *(ii)* the Italian Ministry for Universities and Research, *(iii)* the Luxembourg National Research Fund (G.A. INTER/CHIST/19/14589586), *(iv)* the Scientific and Research Council of Turkey (TÜBİTAK, G.A. 120N680).

References

1. Alexandra, V.A., Badica, C.: Recommender systems: an explainable AI perspective. In: 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1–6. IEEE (08 2021)

2. Anjomshoae, S., Najjar, A., Calvaresi, D., Främling, K.: Explainable agents and robots: results from a systematic literature review. In: AAMAS, Montreal, Canada, May 13–17, pp. 1078–1088 (2019)
3. Aydoğan, R., Festen, D., Hindriks, K.V., Jonker, C.M.: Alternating offers protocols for multilateral negotiation. In: Fujita, K., et al. (eds.) Modern Approaches to Agent-based Complex Automated Negotiation. SCI, vol. 674, pp. 153–167. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-51563-2_10
4. Ayub, R., Ghazanfar, M.A., Maqsood, M., Saleem, A.: A jaccard base similarity measure to improve performance of CF based recommender systems, pp. 1–6 (01 2018). https://doi.org/10.1109/ICOIN.2018.8343073
5. Chen, M., Jia, X., Gorbonos, E., Hoang, C.T., Yu, X., Liu, Y.: Eating healthier: exploring nutrition information for healthier recipe recommendation. Inf. Process. Manage. **57**(6), 102051 (2020)
6. Chi, Y.L., Chen, T.Y., Tsai, W.T.: A chronic disease dietary consultation system using owl-based ontologies and semantic rules. J. Biomed. Inform. **53**, 208–219 (2015)
7. Corrado, S., Luzzani, G., Trevisan, M., Lamastra, L.: Contribution of different life cycle stages to the greenhouse gas emissions associated with three balanced dietary patterns. Sci. Tot. Environ. **660**, 622–630 (2019)
8. Elsweiler, D., Trattner, C., Harvey, M.: Exploiting food choice biases for healthier recipe recommendation. In: Proceedings of the 40th International ACM SIGIR Conference, pp. 575–584. Association for Computing Machinery (2017)
9. Fanda, L., Cid, Y.D., Matusz, P.J., Calvaresi, D.: To pay or not to pay attention: classifying and interpreting visual selective attention frequency features. In: Calvaresi, D., Najjar, A., Winikoff, M., Främling, K. (eds.) EXTRAAMAS 2021. LNCS (LNAI), vol. 12688, pp. 3–17. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-82017-6_1
10. Freyne, J., Berkovsky, S.: Recommending food: reasoning on recipes and ingredients. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 381–386. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13470-8_36
11. Ge, M., Ricci, F., Massimo, D.: Health-aware food recommender system. In: Proceedings of the 9th ACM Conference on Recommender Systems, pp. 333–334. RecSys '15, Association for Computing Machinery, New York, NY, USA (2015)
12. Gibney, M.J., Forde, C.G., Mullally, D., Gibney, E.R.: Ultra-processed foods in human health: a critical appraisal. Am. J. Clin. Nutr. **106**(3), 717–724 (2017)
13. Gunning, D., Aha, D.: Darpa's explainable artificial intelligence (xai) program. AI Mag. **40**(2), 44–58 (2019). Jun
14. Hammond, K.J.: Chef: a model of case-based planning. In: AAAI (1986)
15. Harris, J.A., Benedict, F.G.: A biometric study of human basal metabolism. Proc. Natl. Acad. Sci. **4**(12), 370–373 (1918)
16. Igor, T., Jean, E., Ndamlabin, M., Amro, N., Yazan, M., Stéphane, G.: A decentralized multilevel agent based explainable model for fleet management of remote drones. Procedia Comput. Sci. **203**, 181–188 (2022)
17. Ishizaka, A., Siraj, S.: Are multi-criteria decision-making tools useful? an experimental comparative study of three methods. Eur. J. Oper. Res. **264**(2), 462–471 (2018)
18. Lawo, D., Neifer, T., Esau, M., Stevens, G.: Buying the 'right' thing: designing food recommender systems with critical consumers. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI 2021, Association for Computing Machinery, New York, NY, USA (2021)

19. Lazar, J., Feng, L.H., Hochheiser, H.: Research Methods in Human-Computer Interaction. Willey (2010)
20. Mualla, Y., et al.: The quest of parsimonious XAI: a human-agent architecture for explanation formulation. Artif. Intell. **302**, 103573 (2022)
21. Padhiar, I., Seneviratne, O., Chari, S., Gruen, D., McGuinness, D.L.: Semantic modeling for food recommendation explanations. In: ICDEW, pp. 13–19. IEEE (2021)
22. Popovski, G., Seljak, B., Eftimov, T.: Foodbase corpus: a new resource of annotated food entities. Database J. Biolog. Databases Curation **11**, baz121 (2019)
23. Rubinstein, A.: Perfect equilibrium in a bargaining model. Econometrica: J. Econ. Soc. **50**(1), 97–109 (1982)
24. Samih, A., Adadi, A., Berrada, M.: Towards a knowledge based explainable recommender systems. In: Proceedings of the 4th International Conference on Big Data and Internet of Things. BDIoT2019, Association for Computing Machinery, New York, NY, USA (2019)
25. Soutjis, B.: The new digital face of the consumerist mediator: the case of the 'Yuka' mobile app. J. Cultural Econ. **13**(1), 114–131 (2020)
26. Starke, A., Trattner, C., Bakken, H., Johannessen, M., Solberg, V.: The cholesterol factor: balancing accuracy and health in recipe recommendation through a nutrient-specific metric. In: CEUR Workshop Proceedings, vol. 2959 (2021)
27. Teng, C.Y., Lin, Y.R., Adamic, L.A.: Recipe recommendation using ingredient networks. In: Proceedings of the 4th Annual ACM Web Science Conference, pp. 298–307. WebSci 2012 (2012)
28. Toledo, R.Y., Alzahrani, A.A., Martinez, L.: A food recommender system considering nutritional information and user preferences. IEEE Access **7**, 96695–96711 (2019)
29. Trang Tran, T.N., Atas, M., Felfernig, A., Stettinger, M.: An overview of recommender systems in the healthy food domain. J. Intell. Inf. Syst. **50**(3), 501–526 (2017). https://doi.org/10.1007/s10844-017-0469-0
30. Tran, T.N.T., Felfernig, A., Trattner, C., Holzinger, A.: Recommender systems in the healthcare domain: state-of-the-art and research issues. J. Intell. Inf. Syst. **57**(1), 171–201 (2021)
31. Wang, L., et al.: A dynamic multi-attribute group emergency decision making method considering experts' hesitation. Int. J. Comput. Intell. Syst. **11**(1), 163–182 (2018)