Question Ranking for Food Frequency Questionnaires

Nina Reščič
nina.rescic@ijs.si
Department of Intelligent Systems,
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Mitja Luštrek
mitja.lustrek@ijs.si
Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

ABSTRACT

Food Frequency Questionnaires (FFQs) are probably the most commonly used dietary assessment tools. In the WellCo project, we developed the Extended Short Form Food Frequency Questionnaire (ESFFFQ), integrated in mobile application, in order to monitor quality of users' nutrition. The developed questionnaire returns diet quality scores for eight targets — *fruit intake*, *vegetable intake*, *fish intake*, *salt intake*, *sugar intake*, *fat intake*, *fibre intake* and *protein intake*. This paper explores the single-target problem of question ranking. We compared question ranking of the machine learning algorithms on three different types of features for classification and regression problem. Our findings showed that the addressing problem as a regression problem performs better than treating it as a classification problem and the best performance was achieved by using a Linear Regression on features, where answers were transformed to frequencies of consumption certain food group.

KEYWORDS

nutrition monitoring, FFQs, question ranking

1 INTRODUCTION

Adopting and maintaining a healthy lifestyle has become extremely important and healthy nutrition habits represent a major part in achieving this goal. Self-assessment tools are playing a big role in nutrition monitoring and many applications are including Food Frequency Questionnaires (FFQs) as a monitoring tool, due to they in-expensiveness, simplicity and reasonably good assessment [8, 7]. An FFQ is a questionnaire that asks the respondents about the frequency of consumption of different food items (e.g., "How many times a week do you eat fish?"). In the EU-funded project WellCo we developed and validated an Extended Short Form Frequency questionnaire (ESFFFQ) [4] that was included in a health coaching application for seniors.

In order not to burden the users, we explored the possibilities to get the most information even if one does not answer the whole questionnaire. In our previous work we explored how to find the smallest set of questions that still provides enough information by applying different feature selection techniques [5, 6]. While machine learning has been previously used to detect dietary patterns [2] or estimate nutrient intake [1], at least to our knowledge it has not been used to develop or optimize the FFQs (apart from our previous work). This paper explores the ranking of questions as the next step from selecting the important questions. By ranking the questions by importance and asking them in this order we can expect to predict the best information available at that very moment. We explored the question as a single-target problem for classification and regression problems. Additionally, we tested the algorithms on different representations of features for both type of problem. The findings of this paper could be used for setting the baseline for our future research.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

2 METHODOLOGY

2.1 Problem outline

In our previous research [5, 6] we tried to find subsets of questions that would allow us to ask the users about their dietary habits with as few questions as possible and still get sufficient information to evaluate their nutrition. For this we used the Extended Short Form Food Frequency Questionnaire (ESFFFQ) [4]. The questionnaire returns diet quality scores scores for *fruit intake*, *vegetable intake*, *fish intake*, *salt intake*, *sugar intake*,*fat intake*, *fibre intake* and *protein intake*. We calculate the nutrient intake amounts and from there we further calculate the diet quality scores.

The questionnaire was included in a mobile application, where the system asked the users about their diet with one or two questions per day. The answers were saved into a database and every fortnight the quality scores were recalculated. As it could happen that the users did not answer all the questions by the time the recalculation was done, it was of great importance to ask the questions in the right order. In the terminology of machine learning this would be a feature ranking problem. We explored the problem as a set of single-target problems — separately for individual outcome scores. As three of the diet quality scores (*fruit*, *vegetable* and *fish intake*) are only dependent on one or two questions, the problem of feature ranking is trivial. Therefore we explored the problem for the remaining five targets — *fat intake*, *sugar intake*, *fibre intake*, *protein intake* and *salt intake*.

2.2 Dataset

We got the answers to ESFFFQ from 92 adults as a part of the WellCo project and additionally from 1039 adults included in SIMenu, the Slovenian EUMenu research project [3]. The questions included in the ESFFFQ were a subset of the questions in the FFQ in SIMenu. Furthermore, the answers (consumption frequencies) were equivalent in both questionnaires, and consequently extracting the answers from SIMenu and adding them to the answers from the ESFFFQ was a very straightforward task.

2.3 Feature ranking

To do the experiments, we first randomly split the data into validation and training sets in ratio 1:3. To train the models and rank the features we then used 4-fold cross-validation on the training set and used the average feature importance from all 4 folds as the final feature ranking.

The ranked features were used to predict quality scores (classification problem) and nutrient amount (regression problem), by adding the question as they were ranked. In this paper we present the results for two commonly used machine learning algorithms — Logistic/Linear Regression and Random Forest Classifier/Regressor. To rank the features we used the absolute value of the coefficients in the Linear/Logistic Regression and the `feature_importance` attribute as implemented in the Random Forest Classifier/Regressor in the `sklearn` library.

Additionally we compared different feature representations — features where answers are represented with nominal discrete equidistant values (once per week is represented as integer 2), features where answers were transformed into frequencies of consumption (once per week is represented as approx. 0.14 per day) and features where answers were transformed into amounts of nutrients (once per week is represented as grams/day). In the last represenation, the features differed between the targets *sugar*, *fat*, *salt*, *fibre* and *protein*. We ran the experiments for five diet categories (*fat intake*, *sugar intake*, *fibre intake*, *protein intake* and *salt intake*) for both classification and regression problem. In both cases we started with the best ranked question, trained the model and compared results on train and validation sets. Then we added the second best ranked question, trained the models and compared the results. We added the questions one by one until the last one.

3 RESULTS

3.1 Classification problem

For classification we tried to predict the quality scores for each of the five nutrition categories. There were three scores - 2 (good), 1 (medium) and 0 (bad). The distribution of the scores for all the categories is shown in Table 1.

Table 1: Distribution of target values for classification

Score	Fat	Sugar	Fibre	Protein	Salt
2	51%	74%	26%	79%	32%
1	31%	14%	22%	13%	47%
0	18%	12%	52%	8%	21%

We compared Random Forest Classifier and Logistic Regression for three different types of features - discrete equidistant answers, answers transformed to frequencies and answers transformed to amounts.

Fat. For Random Forest (RF) there was not a big difference between the three representations of the features. With all three, the highest accuracy on the validation set (79%) is achieved with 5 questions and afterwards the accuracy starts falling and stays on the interval between 75% and 79%. This clearly indicates overfitting, which is confirmed by the fact that the accuracy for RF on the training set was 100% from the fifth question. A similar situation happened for all the remaining targets and will not be repeated in the following subsections. On the training set Logistic Regression (LR) had worse results than the RF and it also performed the worst from all algorithms when run on the discrete features. However, when the features are transformed into frequencies or amounts, we get better results on the validations set than with RF.

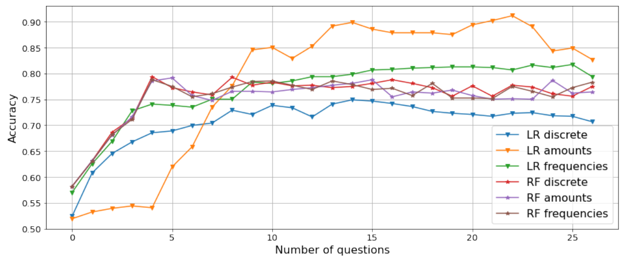

Figure 1: Results on validation set for *fat intake*

Sugar. For *sugar intake* the story is very similar. RF performed fairly well for the first few questions and then the accuracy began to fall. The best performing algorithm was the LR on the features (Figure 2, where the answers were transformed into frequencies.

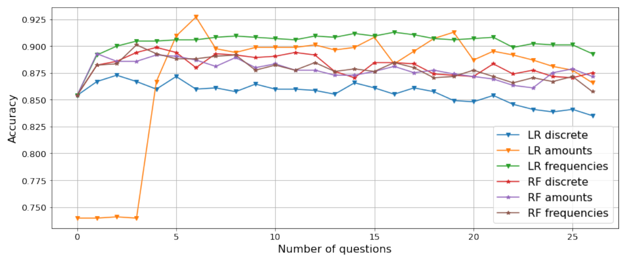

Figure 2: Results on validation set for *sugar intake*

Fibre. For *fibre intake* the RF algorithms performed better for a very long time (Figure 3) and it reached the best accuracy after 6 questions. The LR performed worse, and it did similarly badly on the training set as well.

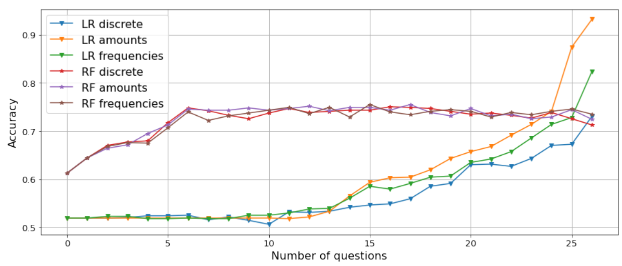

Figure 3: Results on validation set for *fibre intake*

Protein. For *protein intake* (Figure 4) the results are similar to those for *fibre intake*. However, in case of *protein intake* the majority class is 79% and most of the algorithms almost never exceeded this value.

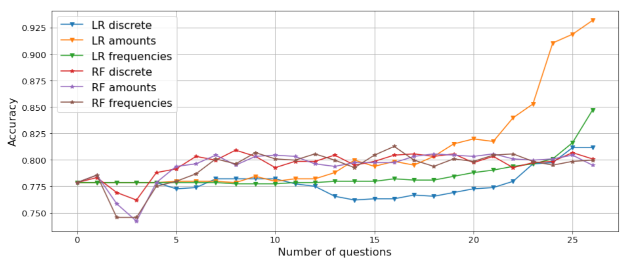

Figure 4: Results on validation set for *protein intake*

Salt. For *salt intake* the best model is the LR on the answers transformed to amounts. As seen in Figure 5, it exceeded the RF algorithms for almost 20% from eleventh added question on and predicted the quality scores with more than 90% accuracy with only 14 questions, which is half of the questionnaire.

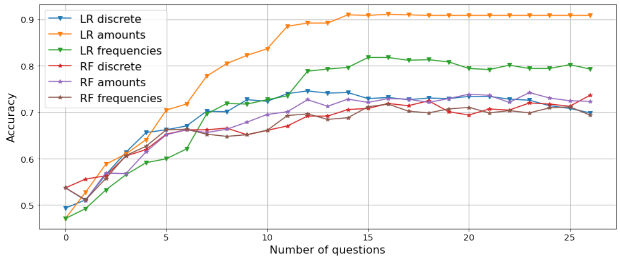

Figure 5: Results on validation set for *salt intake*

3.2 Regression problem

While knowing the quality score is a valid first information whether one's diet is good or not, generally more interesting information is how good (or how bad) it really is. Therefore it is reasonable to look at the same problem as a regression problem , where we try to predict the actual amount (in grams) of consumed nutrients. Again we explored the performance of Random Forest Regressor (RF) and Linear Regression (LR) on the three previously described feature sets.

Table 2: Nutrient intake in grams/day to quality scores

Score	Fat[g]	Sugar[g]	Fibre[g]	Protein[g]	Salt[g]
2	≤ 74	≤ 55	≥ 30	≥ 55	≤ 6
1	else	else	else	else	else
0	≥ 111	≥ 82	≤ 25	≤ 45	≥ 9

Fat. The best performing algorithm for *fat intake* was the LR on the answers transformed to frequencies. The overfitting of the RF is even more visible than with the classification problem as the errors for these models did not fall under 20 grams even if all the questions were used, while the error of the LR on the feature sets where the answers are transformed to frequencies or amounts was smaller than 5 grams from eleven included questions (Figure 6).

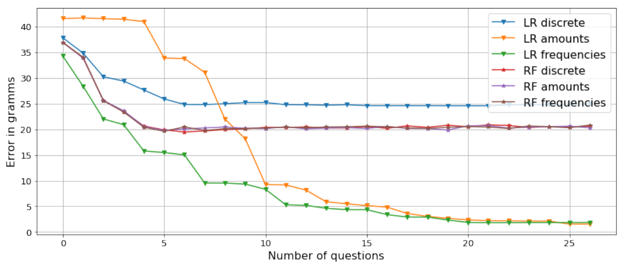

Figure 6: Results on validation set for *fat intake*

Sugar. Similarly to *fat intake*, LR with the 'frequency features' performed best (Figure 7). However the LR on the 'amounts features' performed well for more than 15 questions, but predicted the worst for the first eleven included questions.

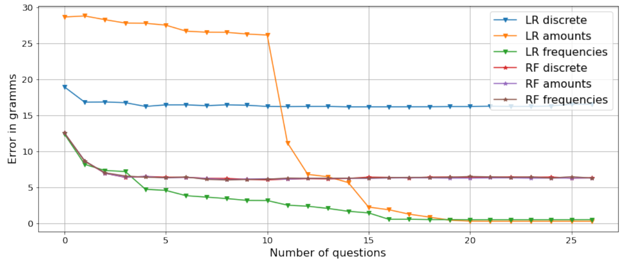

Figure 7: Results on validation set for *sugar intake*

Fibre. Classification for *fibre intake* was very bad, however, when considering it as a regression problem, the LR on 'frequency' features' predicted the amounts with error smaller than 2 grams when more than eleven questions were used 8. Considering Table 2 this means that predicting how bad/good the *fibre intake* was done better then predicting if it is bad or good.

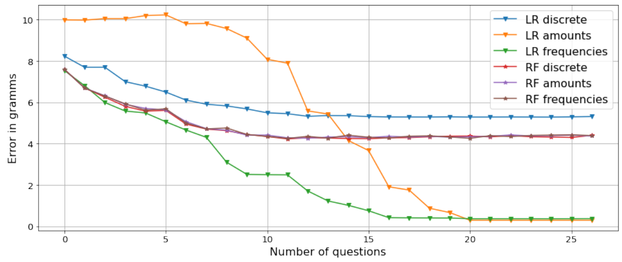

Figure 8: Results on validation set for *fibre intake*

Protein. For *protein intake* all algorithms had a similar performance up to ten included questions, however, the LR on the 'frequency features' started to perform better and better with each added questions and predicted the amount of protein consumption with error of 5 grams (Figure 9).

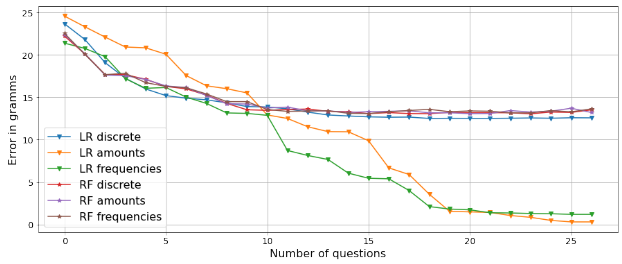

Figure 9: Results on validation set for *protein intake*

Salt. Similarly to the *protein intake* all algorithms performed with a comparable error up to nine included questions, and after that LR using the features transformed to frequencies started to perform way better and predicted *salt intake* with error smaller than 1 gram with eleven included questions (Figure 10).

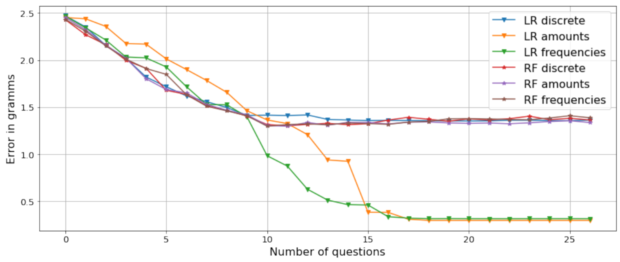

Figure 10: Results on validation set for *salt intake*

3.3 Discussion

We compared performance of feature ranking for two different machine learning algorithms on three different types of features for both classification and regression problems. While the classification problem might give the general idea about one's dietary habits, it is inclined towards overfitting even for very simple models, such as Logistic Regression, while more complex algorithms, Random Forest Classifier in our case, are even more subject to this deficiency. By predicting amounts instead of quality scores, one gets information about how good/bad the dietary habits are instead of just if they are good or bad.

Transforming features from discrete equidistant values to frequencies or amounts of nutrients proved to be a very good approach. The transformation gave better results for both classification and regression problem for both Random Forest Regressor/Classifier and Logistic/Linear Regression. While the performance of both algorithms on features transformed to frequencies and features transformed to amounts for the classification problem was comparable, and Linear Regression on features transformed to amounts gave markedly better results for *salt intake*, the Linear Regression on features transformed to frequencies outperformed all other combinations of features and algorithms for the regression problem for all of the targets. The reason for this is that linear regression on amounts is a very good match in the sense that the target variable (total amount) is the sum of all features (partial amounts).

Transforming the features to frequencies instead to amounts has another advantage — frequencies transformed to amounts are specific to each target, while features transformed to frequencies are equal for all targets. This is an important finding for possible future research where one would address ranking of questions as a multi-target problem. Additionally, regression problem using Linear Regression on features transformed to frequencies could solve as a baseline for future experiments.

4 CONCLUSION AND FUTURE WORK

Ranking the questions of FFQs when it could be expected that not all of the questions will be answered is an important step when building models for predicting quality of one's diet. In this paper we compared two feature ranking algorithms on three different types of features for classification and regression problem for five targets. The findings of this paper show that considering the problem as a regression problem on features transformed to frequencies and using a simple machine learning algorithms (Linear Regression) gives the best results for all five targets and provides baseline for future experiments.

There are several possibilities for future work. As hinted in the previous section, the question of multi-target question ranking is one of the first that appears — one might want to monitor several nutrition quality scores but still would want to avoid answering too many questions. Next, probably more important and interesting research problem, is how to use the answers already provided to our advantage — so instead of statically ranking the questions we would rather explore how we could improve the prediction performance by dynamically ranking and asking the questions.

ACKNOWLEDGMENTS

WellCo Project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769765.

REFERENCES

[1] Elizabeth L. Chin, Gabriel Simmons, Yasmine Y. Bouzid, Annie Kan, Dustin J. Burnett, Ilias Tagkopoulos, and Danielle G. Lemay. 2019. Nutrient estimation from 24-hour food recalls using machine learning and database mapping: a case study with lactose. *Nutrients*, 11, 12. ISSN: 2072-6643. DOI: 10.3390/nu11123045.

[2] M Gjoreski, S Kochev, and Reščič, N and Gregorič, M. and Eftimov, T. and Koroušić Seljak, B. 2019. Exploring dietary intake data collected by fpq using unsupervised learning. *2019 IEEE International Conference on Big Data*, 5126–5130. DOI: 10.1109/BigData47090.2019.9005715.

[3] Matej Gregorič, Urška Blaznik, Nataša Delfar, Metka Zaletel, Darja Lavtar, Barbara Koroušić-Seljak, Petra Golja, Katja Zdešar Kotnik, Igor Pravst, Nataša Fidler Mis, Stojan Kostanjevec, Majda Pajnkihar, Tamara Poklar Vatovec, and Ada Hočevar-Grom. 2019. Slovenian national food consumption survey in adolescents, adults and elderly : external scientific report. *EFSA Supporting Publications*, 16, 11, 1729E. DOI: 10.2903/sp.efsa.2019.EN-1729.

[4] Nina Reščič, Eva Valenčič, Enej Mlinarič, Barbara Koroušić Seljak, and Mitja Luštrek. 2019. Mobile nutrition monitoring for well-being. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, London, United Kingdom, 1194–1197. ISBN: 9781450368698. DOI: 10.1145/3341162.3347076. https://doi.org/10.1145/3341162.3347076.

[5] Nina Reščič, Tome Eftimov, Barbara Koroušić Seljak, and Mitja Luštrek. 2020. Optimising an ffq using a machine learning pipeline to teach an efficient nutrient intake predictive model. *Nutrients*, 12, 12. ISSN: 2072-6643. DOI: 10.3390/nu12123789. https://www.mdpi.com/2072-6643/12/12/3789.

[6] Nina Reščič, Tome Eftimov, and Barbara Koroušić Seljak. 2020. Comparison of feature selection algorithms for minimization of target specific ffqs. In *2020 IEEE International Conference on Big Data (Big Data)*, 3592–3595. DOI: 10.1109/BigData50022.2020.9378246.

[7] JS Shim, K Oh, and HC Kim. 2014. Dietary assessment methods in epidemiologic studies. *Epidemiol Health*, 36. DOI: 10.4178/epih/e2014009.

[8] Frances Thompson and T Byers. 1994. Dietary assessment resource manual. *The Journal of nutrition*, 124, (December 1994), 2245S–2317S. DOI: 10.1093/jn/124.suppl_11.2245s.