



Diet Modelling: Combining Mathematical Programming Models with Data-Driven Methods

Ante Ivancic^(✉), Argyris Kanellopoulos, and Johanna M. Geleijnse

Wageningen University and Research, Wageningen, The Netherlands
{ante.ivancic, argyris.kanellopoulos, marianne.geleijnse}@wur.nl

Abstract. Mathematical programming has been the principal workhorse behind most diet models since the 1940s. As a predominantly hypothesis-driven modelling paradigm, its structure is mostly defined by a priori information, i.e. expert knowledge. In this paper we consider two machine learning paradigms, and three instances thereof that could help leverage the readily available data and derive valuable insights for modelling healthier, and acceptable human diets.

Keywords: Machine learning · Diet modelling · Consumer preferences

1 Introduction

Current societies are confronted with major challenges. The confluence of population, economic development, and environmental pressures resulting from globalization and industrialization reveals an increasingly resource-constrained world in which predictions point to the need to do more with less and in a more efficient way [1].

Although global food production of calories has kept pace with population growth, more than 820 million people have insufficient food and many more consume low-quality diets that cause micronutrient deficiencies and contribute to a substantial rise in the incidence of diet-related obesity and diet-related non-communicable diseases [2], such as cardiovascular disease, type 2 diabetes, and various types of cancer. As a matter of fact, most NCDs have a root cause in an unhealthy diet [3, 4]; a diet that does not fulfill energy and nutrient requirements for healthy growing and aging [5].

The concept of sustainable diets presents an opportunity to successfully advance commitments to sustainable development and the elimination of poverty, food and nutrition insecurity, and poor health outcomes [1]. Sustainable diets can be defined as those with low environmental impacts that contribute to food and nutrition security and to healthy life for present and future generations [6].

Designing and promoting sustainable diets is a complex task. Mathematical and computational models that can capture the complexity of the problem, and

devise a sustainable nutritional strategy are needed [7]. Moreover, such models can help to set priorities for interventions/policy measure, which might result in a more sustainable consumption pattern that would act as a driver of sustainable production, since current diet and production patterns are among the most important drivers of environmental pressure [8].

Capturing diet modelling complexity is a twofold challenge. Firstly, diet modellers gain their knowledge over the course of many years of education and professional experience, ultimately constructing an immensely convoluted model of what and when a specific diet can or can not be considered healthy. Translating such an intricate biological knowledge system into a set of computer instructions is a nontrivial challenge. Secondly, the goal of diet models is to change – or at least strive towards changing – consumers’ dietary patterns, for which they need, at least to some extent, appeal to their preferences. Consumers rarely explicitly state their preferences, let alone explain the reasons behind them. However, in today’s data-loaded world, their actions (e.g. recorded supermarket transactions) speak for themselves. The necessity of turning data into actionable insights bodes ill for current diet models, given their inability to do so.

Machine learning, a sub-field of artificial intelligence specialized for automated pattern recognition/feature extraction/(lossy) data compression [9], is a likely candidate for providing algorithmic solutions current diet models could benefit from, namely because of their ability to reshape vast amounts of data into useful information.

In this paper, we consider two machine learning paradigms, and three instances thereof for which we argue to be valuable additions to current diet models. We focus on the two aspects of diet modelling, namely nutrient importance weighting and consumer preferences. For nutrient importance weighting, where the goal is to assign importance weight to every nutrient so as to estimate the overall diet health score, we consider supervised learning approach (binary classification), and suggest two computational methods that could facilitate the translation of expert knowledge into a set of correlated importance weights - the *Principle of maximum entropy* and *gradient boosted decision trees*. For estimating consumer preferences, we consider the concept of recommendation systems, and suggest one possible instance thereof, namely the *Top-N recommendation system based on mutual information and entropy weighting*.

2 Current Diet Modelling Paradigms

The concept of diet modelling dates back to at least 1940s, when the American economist and Nobel laureate Georges Stigler utilized mathematical programming (mathematical optimization) to solve the “diet problem” - a problem of finding the least costly combination of food items that satisfies all nutrient requirements [10].

Mathematical programming remains the principal workhorse behind the majority of today’s diet models and can be characterized by the three main components: decision variables, an objective function, and a set of constraints [11].

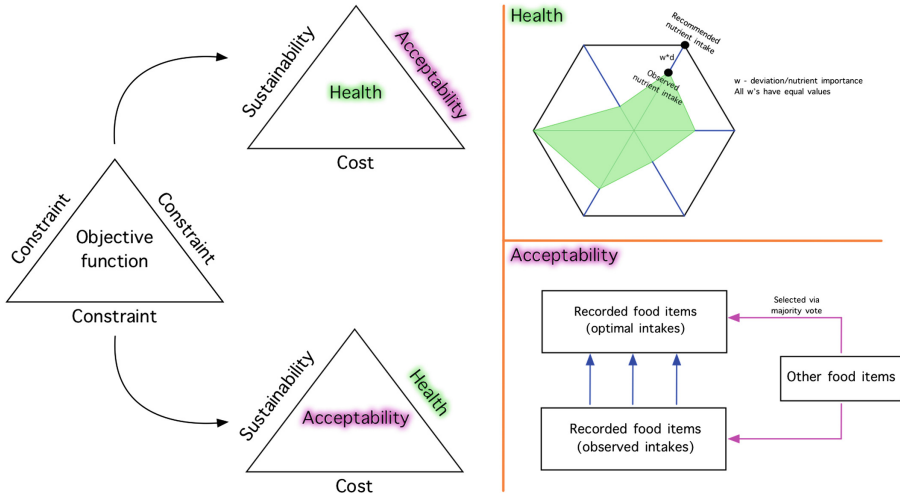


Fig. 1. The majority of diet models are based on some form of mathematical programming, with the two most common objectives: health and acceptability.

It also comes in many forms, depending on the types of decision variables (integers, reals), and the functional form of the objective function. Some of the most prevalent mathematical programming paradigms for diet modelling are linear programming, quadratic programming, mixed-integer programming, and goal-programming [12] (Fig. 1).

During the optimization process, the mathematical programming model aims to search for values for the decision variables that optimize (maximize or minimize) the objective function, while adhering to the preset constraints. In the context of nutrition, diet models based on mathematical programming aim to find the set of food item quantities that optimize a specific objective, (e.g. total diet cost, environmental objective, total deviation from an observed diet), while satisfying constraints such as nutritional recommendations, total energy intake, etc. [12].

2.1 Nutrient Importance and Consumer Preferences

Diet models based on mathematical programming can be classified as hypothesis-driven methods [13], meaning that most of their components are selected based on a priori information. That is, the modeller tries to translate his/her expert knowledge into an appropriate model structure.

For instance, in case of optimizing diet healthiness where the goal is to minimize deviations between observed nutrient intake values and their respective recommended targets, each deviation will be assigned a specific weight, reflecting its relative importance. A total sum of weighted deviations is then equivalent to the overall diet health score, i.e. diet healthiness.

Selecting a set of weights that accurately reflect reality is a challenging task, while almost all diet modellers resort to a uniform set of weights, regardless of the setting [14,15]. We argue that it would be beneficial for diet modellers to have a way of adjusting the weights in case of their disagreement with the uniform set.

Furthermore, in order to generate acceptable optimized diets, diet models have to somehow be able to model the concept of “acceptability”, i.e. somehow model preferences of each consumer. This is currently done by taking into account food consumption distribution in the constraints, or by deriving the average observed diet if the aim is to stay as close as possible to current dietary habits [12]. The former approach generally takes a lot of hard-coding, rules of thumb, and has to be re-implemented manually every time the setting changes, whereas the latter might be “too personalized”, in the sense that the diet model puts a lot of focus on the observed food items, while disregarding the rest. Moreover, it has been observed that in order to meet nutritional targets, observed individual diets generally have to be expanded with new food items, which tends to be done in an arbitrary fashion [12]. Having a more generic, possibly data-driven approach that can collect and process consumer data from multiple sources, and subsequently deliver personalized food item recommendations, could add a lot of value to the already existing diet models, and possibly bridge the gap between hypothesis-driven and data-driven methods.

3 Data-Driven Approaches to Diet Modelling

3.1 Inferring Function Weights

As described in the previous section, while modelling diets via deviations from the optimal diet, objective function weights represent the “importance” of each deviation. Currently, most diet models resort to a uniform set of weights, regardless of the setting (e.g. different sub-populations and their health weights). Indeed, the weights can be changed manually, however given the multidimensional nature of the problem, it is arguably very hard for the modeller to accurately translate its own beliefs into a set of correlated weights.

We consider two possible methods for dealing with this challenging task, namely the *Principle of maximum entropy (MaxEnt)*, and *Gradient boosted decision trees*. MaxEnt is relatively easy to implement, converges to an exact solution (convex optimization), and is highly usable, because once implemented, it requires almost no prerequisite knowledge or technical skills, apart from nutritional expertise. On the other hand, gradient boosted decision trees tend to be exceptionally successful while dealing with tabular data, giving out state-of-the-art results on many standard classification benchmarks [16]. Both methods are explained in the subsequent sections.

Principle of Maximum Entropy. MaxEnt is a general method for estimating probability distributions from data. The core principle behind MaxEnt is

that when nothing is known, the distribution should be as uniform as possible, that is, have maximal information entropy. As bits of information are becoming available, the distribution is updated in a way that adheres to the constraints imposed by the new information, while maximizing the entropy [17].

Essentially, where current models end, MaxEnt begins. That is, starting with a uniform set of weights, diet modellers are able to interact with MaxEnt via pairwise comparisons of diets, i.e. they are presented with diet pairs, and are instructed to select a “better” (e.g. healthier, more sustainable, etc.) diet, according to their expert knowledge. Their classification results get translated into a set of inequalities which will constrain MaxEnt’s search space, in order to find a set of weights that are in line with the expert knowledge.

Gradient Boosted Decision Trees. Decision tree learning is a type of predictive modelling approach commonly used within the machine learning/statistics domain, both for regression and classification problems. In order to make predictions, decision trees stratify the predictor space into a number of simple regions, while the splitting rules used for stratification can be represented as a directed acyclic graph whose child nodes can only have a single parent, i.e. a tree, hence their name. In their basic form, they tend not to be as accurate as some other prediction models, however their prediction accuracy can be increased significantly if coupled with some other machine learning methods, such as gradient boosting [16, 18]. Furthermore, decision trees tend to be highly interpretable (e.g. each decision can easily be manually inspected), robust (e.g. do not require significant data preprocessing), scalable, and suitable for parallel and distributed computation [18].

Because of their interpretability and high predictive power, we consider a decision tree-based method as a viable option for inferring objective function weights, namely *gradient boosted decision trees*. We cast the problem of inferring function weights as a binary classification problem, where the setting is the same as with the maximum entropy principle - diet modellers are presented with diet pairs and are instructed to select the one they prefer more, with respect to the objective function. By doing so, they generate input-output pairs on which gradient boosted decision trees are then trained/validated. Upon convergence, we select the best performing decision tree, and rank input features according to their importance, which in this case is synonymous with their *total information gain* [18]. We further normalize the results, so as to obtain a proper probability distribution.

Besides inferring objective function weights, the method can provide diet modellers with a variety of other information. For instance, features that appear together in a traversal path are interacting with one another (since the condition of a child node is predicated on the condition of the parent node), which gives modellers the ability to discover interactions among features [16].

3.2 Inferring Consumer Preferences

Diet acceptability modelling has been deemed interesting and important both as a constraint, and an objective, depending on the modelling approach. It is generally modelled as the total deviation between the optimal food item intakes, and the food item intakes of the current diet. The assumption is that if two diets do not deviate significantly in terms of food items and their quantities, they can be considered similar, and almost equally acceptable. This definition of acceptability implicitly takes into account a variety of different factors that might contribute to the diversity of dietary patterns, for instance cultural and lifestyle differences [12].

However, modelling diet acceptability solely through total deviation leaves a lot to be desired. For instance, such approach does not provide the option to search for likely preferable food items that have not been recorded in a consumer's diet, which puts a severe limitation on diet model flexibility. As mentioned before, in order to meet nutritional targets, observed individual diets generally have to be expanded with new food items, which, if done in an arbitrary fashion, could have detrimental effects on diet acceptability. The same thing applies in case a modeller would like to diversify the observed diets by including new food items.

In the following section, we consider a data-driven modelling paradigm based on the concept of *recommendation systems* [19], namely the *Top-N recommendation system based on mutual information and entropy weighting*, that can analyze consumer historical data, and turn it into actionable insights that can complement the already existing diet models. By doing so, we are establishing one of potentially many links between the hypothesis-driven methods, and more data-driven machine learning methods.

Top-N Recommendation System Based on Mutual Information and Entropy Weighting. In order to meet all nutritional constraints, consumers' diets often have to be expanded with new, previously unobserved food items. Most current diet models base their selection of new food items on some form of the majority vote, that is selecting those unobserved foods that are present in e.g. $\geq 50\%$ of diets in the sample [12]. Although better than random selection, such an approach leaves enough room for improvement.

We consider a recommendation system that can process user data from potentially multiple sources, and leverage algorithmically derived information on consumers' routine behavioral patterns so as to subsequently deliver personalized product recommendations. The recommendation system consists of two computational steps - the preprocessing and pairwise similarity computation.

Preprocessing of the input data can significantly facilitate the extraction of information [20]. It comes in many flavors, ranging from very simple procedures such as normalization and standardization, various kernel functions, weighting functions [21], (non)linear dimensionality reduction methods, to more sophisticated and automatic feature extractors such as deep neural networks [20]. The selection of one or more preprocessing steps will depend on a variety of factors,

including, but not limited to, available computing power, and the amount and type of data. Given that the majority of consumer behavior data is not publicly available, diet modellers predominantly leverage relatively scarce data sets obtained via questionnaires [12], hence why we consider a handcrafted weighting function based on the concept of information entropy, which has been shown to perform significantly better than most handcrafted weighting functions in some tasks [22].

As the name suggests, similarity computation step serves for computing either pairwise food item or consumer similarities. Selecting an appropriate similarity measure is of crucial importance. We consider the mutual information, a core information-theoretic quantity that acts as a general measure of dependence between two random variables [23], as opposed to some other commonly employed correlation metrics (e.g. Pearson correlation coefficient), which measure only linear dependence.

Similarity computation ultimately results in a fully connected similarity graph that can be queried in many ways. In case of food item similarities, such a graph can for instance support the prediction of top N likely preferable food items by consumers that have not reported consuming those food items.

For each consumer m_i that has purchased a set U of food items, we compute the set C by taking the union of the k most similar food items for each item $n_i \in U$. After that, we remove all food items from C that are already in U . Then, for each item $c \in C$ we compute its similarity to the set U by summing the similarities between all food items $n_i \in U$ and c , using only the k most similar food items of n_i . Lastly, the food items in C are sorted in non-increasing order with respect to their similarity to the set U , and the first N food items are selected for the recommendation. Clearly, N and k are tunable parameters, and their selection can affect the speed and quality of recommendations [24].

4 Conclusion

Diet models based on mathematical programming have been used extensively during the last couple of decades, and have stood the test of time. With straightforward structure, fast execution, and high usability, mathematical programming poses as an obvious first choice for diet modelling. However, being mainly hypothesis-driven, such diet models often neglect important aspects of today’s world - the abundance of user data, and the availability of algorithms that can turn data into actionable insights.

In this paper, we provide just a few examples of available data-driven methods that could greatly facilitate the diet modelling process. With MaxEnt and Gradient boosted decision trees, we provide diet modellers with the means for interacting with their diet models, so as to translate their expert knowledge into a “machine-readable” format. With the Top-N recommendation system, we are enriching the existing diet models with algorithmic and much more information-rich “word-of-mouth” recommendations.

Indeed, data-driven algorithms come with a few caveats. For instance, current state-of-the-art methods are still rather data-inefficient, meaning that they need

a significant amount of data to obtain high generalization power. Furthermore, they can also be energy-inefficient, in the sense that large amount of data require significant computational power to be processed in a reasonable amount of time.

References

1. Johnston, J.L., Fanzo, J.C., Cogill, B.: Understanding sustainable diets: a descriptive analysis of the determinants and processes that influence diets and their impact on health, food security, and environmental sustainability. *Adv. Nutr.* **5**(4), 418–429 (2014)
2. Willet, W., et al.: Food in the anthropocene: the EAT-lancet commission on healthy diets from sustainable food systems. *Lancet* **393**(10170), 447–492 (2019)
3. Lassale, C., Gunter, M.J., Romaguera, D., Peelen, L.M., Van der Schouw, Y.T., Beulens, J.W.J., et al.: Diet quality scores and prediction of all-cause, cardiovascular and cancer mortality in a pan-European cohort study. *PLoS ONE* **11**(7) (2016)
4. Miniñane, A.M., et al.: Low-grade inflammation, diet composition and health: current research evidence and its translation. *Br. J. Nutr.* **114**(07), 999–1012 (2015)
5. Fanzo, J.: Ethical issues for human nutrition in the context of global food security and sustainable development. *Glob. Food Secur.* **7**, 15–23 (2015)
6. Macdiarmid, J.I., et al.: Sustainable diets for the future: can we contribute to reducing greenhouse gas emissions by eating a healthy diet? *Am. J. Clin. Nutr.* **96**(3), 632–639 (2012)
7. Irz, X., Leroy, P., Requillart, V., Soler, L.: Beyond wishful thinking: integrating consumer preferences in the assessment of dietary recommendations. *PLoS ONE* **11**(6) (2016)
8. Reisch, L., Eberle, U., Lorek, S.: Sustainable food consumption: an overview of contemporary issues and policies. *Sustain. Sci. Pract. Policy* **9**(2), 7–25 (2016)
9. Schmidhuber, J.: Deep learning in neural networks: an overview, pp. 7–25. [arXiv:1404.7828](https://arxiv.org/abs/1404.7828) (2014)
10. Stigler, G.J.: The cost of subsistence. *J. Farm Econ.* **27**(2), 303–314 (1945)
11. Danzig, G.B.: *Linear Programming and Extensions*. Princeton University Press, Princeton (1963)
12. Gazan, R., Brouzes, C.M.C., Vieux, F., Maillot, M., Lluch, A., Darmon, N.: Mathematical optimization to explore tomorrow’s sustainable diets: a narrative review. *Adv. Nutr.* **9**(5), 602–626 (2018)
13. Previdelli, A.N., de Andrade, S.C., Fisberg, R.M., Marchioni, D.M.: Using two different approaches to assess dietary patterns: hypothesis-driven and data-driven analysis. *Nutrients* **8**(10), 593 (2016)
14. Tyszler, M., Kramer, G., Blonk, H.: Just eating healthier is not enough: studying the environmental impact of different diet scenarios for Dutch women (31–50 years old) by linear programming. *Int. J. Life Cycle Assess.* **21**(5), 701–709 (2016)
15. van Dooren, C., Tyszler, M., Kramer, G.F.H., Aiking, H.: Combining low price, low climate impact and high nutritional value in one shopping basket through diet optimization by linear programming. *Sustainability* **7**(9), 12837–12855 (2015)
16. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, pp. 785–794 (2016)

17. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**(4), 620 (1957)
18. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer, New York (2001). <https://doi.org/10.1007/978-0-387-21606-5>
19. Lu, L., Medo, M., Yeung, C.H., Zhang, Y.-C., Zhang, Z.-K., Zhou, T.: Recommender systems. *Phys. Rep.* **519**(1), 1–49 (2012)
20. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
21. Manning, C.D., Raghavan, P.: *Introduction to Information Retrieval*. Cambridge University Press, New York (2008)
22. Nakov, P., Popova, A., Mateev, P.: Weight functions impact on LSA performance. In: *Proceedings of the EuroConference Recent Advances in Natural Language Processing (RANLP 2001)*, pp. 187–193. (2001)
23. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edn. Wiley, New York (2006)
24. Karypis, G.: Evaluation of item-based top-N recommendation algorithms. In: *9th Proceedings of the Tenth International Conference on Information and Knowledge Management*, pp. 247–254. ACM, New York (2001)