Towards modelling SHARP diets, based on nutritional adequacy, sustainability metrics and population diversity parameters

Deliverable No. D7.4

Ante Ivancic, Argyris Kanellopoulos, Jacqueline Bloemhof-Ruwaard, Marianne Geleijnse (WU)

SUSFANS DELIVERABLES

Version	Release date	Changed	Status	Distribution
V1	30/10/2018	-	Final	Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 633692

SUSFANS Deliverable document information

Project name	SUSFANS
Project title:	Towards modelling SHARP diets, based on nutritional adequacy, sustainability metrics and population diversity parameters
Project no	633692
Start date:	April 2015
Report:	D7.4
Work package	WP7
WP title (acronym):	Modelling SHARP diets for EU-consumers
WP leader:	WUR, Marianne Geleijnse
Period, year:	2018
Responsible Authors:	Ante Ivancic (WUR), Argyris Kanellopoulos (WUR), Marianne Geleijnse (WUR), Jacqueline Bloemhof-Ruwaard (WUR)
Participant acronyms:	WUR
Dissemination level:	Public
Version	V1
Release Date	30/10/2018
Planned delivery date:	30/10/2018
Status	Final
Distribution	Public

Dissemination level of this report

Public

ACKNOWLEDGMENT & DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 633692. Neither the European Commission nor any person acting on behalf of the Commission is responsible for how the following information is used. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

TABLE OF CONTENT

DELIVERABLE SHORT SUMMARY FOR USE IN MEDIA

Data Envelopment Analysis-based diet model serves as an elegant tool for optimizing diets in an innovative way. Its generic structure allows for a variety of extensions to be made on top of the core model. Its common primary objective is diet health optimization, while auxiliary objectives can be diverse, for instance diet sustainability and preferability. If the appropriate diet sustainability metric is available, its inclusion in the model is relatively straightforward, whereas diet preferability presents a much more significant challenge. Instead of inferring consumer preferences by computing consumer similarities, as is the case with some data envelopment analysis-based diet models, our approach is based on food item similarities, which allows for much more intuitive and diverse food item recommendations, including those for food items that have not been observed in consumers' diets.

TEASER FOR SOCIAL MEDIA

Diet modelling has been dominated by linear programming models for many years, however their success has been limited, while their inability to extract value from data in our information-driven world has become readily apparent. Increasing consumers' diet healthiness has been the primary task of almost all diet models, however to actually change patterns of consumers' purchasing behavior, models have to learn their preferences, so as to recommend diet alternatives that are both healthy, and appealing. We present our data-driven approach that leverages food item similarities as the main building blocks of diet recommendations, which arguably represents a paradigm shift in the way we optimize diets. Furthermore, we switch from exploiting what is known to be preferable (i.e. what we observe in a consumer's diet), to exploring what is likely preferable, thereby allowing our diet model to "think outside the box".

Keywords: sustainability, preferability, information entropy, mutual information

Extended SHARP model

Ante Ivancic, Wageningen University

The SUSFANS deliverable D7.4 proposes an extension of the basic SHARP model presented in the SUSFANS deliverable D7.3. The basic SHARP model is based on the Data Envelopment Analysis (DEA), a commonly used nonparametric benchmarking method within the operations research domain. The basic model's primary goal is to identify healthier diets within a particular age/gender group, thereby addressing nutritional adequacy and, to some extent, population diversity parameters. Besides computing diet healthiness, the basic model includes a consumer similarity metric, namely the absolute loss function, which serves as a way of estimating consumer preferences. Such approach would arguably result in diet recommendations that are to some extent both culturally acceptable and economically feasible.

However, the absolute loss comes with a few caveats, namely the intrinsic bias towards diet recommendations that are more-or-less the same as the original diet, in terms of food items that they consist of. Stated differently, the recommended diets will rarely include food items that have not been recorded in a particular consumer's purchase history. This poses as a potential drawback, given that it is a well know fact that many food items serve as either substitute for, or complements to one another, which the absolute loss does not take into account. Furthermore, in case there is a need for recommending a certain food item not present in a particular consumer's diet, for whatever reasons, the absolute loss can not estimate the probability of that food item being preferred by the consumer.

Another drawback of the absolute loss-based recommendations lies in the fact that, in case the model needs to provide reasons for its recommendations, its internal workings are not necessarily intuitive to a human modeller. For instance, a consumer A receives a diet recommendation based on the fact that he/she is similar to consumers B, C, and D. It is arguably extremely hard for the modeller to assess the validity of the results.

In the SUSFANS deliverable D7.4 we have tried to address the aforementioned drawbacks, by proposing an alternative approach for computing preferability, which is primarily based on food item similarities. The hypothesis was that similar food items are bought by similar consumers, which would automatically take into account various aspects of human preferences (determined by consumers' culture, socio-economic status, lifestyle, etc.), while at the same time render our approach fairly transparent, and intuitive. For instance, instead of selecting a consumer, and retrieving a list of similar consumers whose quality

is hard to validate, our preferability module constructs a food item graph, and for each food item, it is able to retrieve a list of the most similar food items. Manually inspecting such results is arguably much easier, given that the modeller only has to compare food items, and not consumers (i.e. potentially long lists of food items).

Furthermore, our proposed approach allows for consumers' diets to be enriched with food items that they have not reported in their diets surveys. For instance, if the model predicts a high probability of a certain food item being preferred by a certain consumer, and if that food item has a high sustainability index, such recommendation could lead to a significant increase in diet's sustainability, which might not be achievable had the algorithm been constrained only to the reported set of food items.

The approach is presented in the remainder of this deliverable. The forthcoming deliverable D7.5 will report on the inclusion of sustainability metrics within the DEA approach and on the application of the model on nutrition surveillance data for the four focus countries in SUSFANS (Czech Republic, Denmark, France and Italy) – effectively establishing the SHARP model as a component of the SUSFANS modelling toolbox.

Extended DEA-based diet model

Technical report

Ante Ivancic[1], Argyris Kanellopoulos[1]
Jacqueline Bloemhof-Ruwaard[1], Johanna M. Geleijnse[2]
[1]Operations Research and Logistics Department
[2]Division of Human Nutrition and Health
Wageningen University & Research
The Netherlands

24-09-2018

Abstract

Data Envelopment Analysis-based diet model serves as an elegant model for optimizing diets in an innovative way. Its generic structure allows for a variety of extensions to be made on top of the core model. Diet healthiness is currently the dominant objective of data envelopment analysis-based diet models, while the rest, such as diet preferability and diet sustainability, are either rudimentary, or non-existent. Diet sustainability, in the form of a metric, can be easily included in the current models, while diet preferability arguably needs to be developed as a separate, more elaborate module, that could support the core data envelopment analysis-based diet model with its outputs. Instead of computing consumer preferences solely through the absolute loss metric (as presented in the SUSFANS deliverable 7.3), which tends to be restricted to the set of food items previously purchased by a consumer, we propose an alternative approach that is based on food item similarities, which would allow for much more diverse food item recommendations, i.e. the recommendations will including the food items that have not been recorded in a consumer's diet. Furthermore, incorporating food item similarities into the absolute loss (or any other metric) would arguably result in more realistic consumer similarity scores.

1 Introduction

The intention of this technical report is to summarize, and build on the idea presented in (Kanellopoulos et al., 2018), where the authors developed a diet model based on Data Envelopment Analysis (DEA) (Cooper et al., 2004). In (Kanellopoulos et al., 2018), the model takes in a sample of diets represented by their nutrient content, and for each of them, it searches for a diet (or a linear combination of diets) that is present in the data set, and at least as healthy as the original one - a procedure that has been termed "benchmarking". Diet healthiness refers to the degree of adherence to nutrient intake recommendations (EFSA, 2017), which is computed through the loss function imposed by the model.

Furthermore, in (Kanellopoulos et al., 2018) the authors proposed three different versions of the DEA model, of which one contained an additional search criterion, namely the diet "preferability" - the likelihood of an alternative diet being preferred by a consumer. That is, when the model identifies a region in the search space that contains exclusively healthier diets, it switches its priority

from identifying the healthiest alternative, to identifying the most similar one, where the notion of diet similarity is equated with diet preferability. Just like the healthiness loss function, the similarity loss function is predefined by the modeler, and is, to a certain degree, subjective (Hastie et al., 2009). Moreover, the performance of the loss function is heavily dependent on the choice of data representation (Bengio et al., 2013). In this case, no transformation has been applied to the raw data, apart from standardizing diets to 2000 kcal per day.

Our aim is to increase the proposed DEA model's capacity, in the sense of including an additional criterion that would be taken into account while benchmarking diets, namely the food sustainability metric, which importance has been increasing at an accelerating rate (Macdiarmid et al., 2012). Furthermore, we argue that a separate "preferability module", in the form of a recommendation system with a preprocessing/data transformation layer that results in a representation of the data that can support effective similarity computations (either between consumers, or food items), would serve as a valuable extension to the core DEA model. The details of such a module will be discussed in the section 4 of this report.

In the following section, we will give a brief summary of the DEA diet model developed in (Kanellopoulos et al., 2018), and will identify some of its potential drawbacks, which will serve as the motivational starting points for the extensions that will be presented in the subsequent sections.

2 DEA diet model - a brief summary

Data Envelopment Analysis (DEA) is a commonly used nonparametric benchmarking method within the operations research domain (Cooper et al., 2004). It has been applied in a variety of different fields, such as banking, healthcare, agricultural economics, transportation, and education. In general, the aim of DEA is to identify Decision Making Units (DMUs) that convert inputs (i.e. less-is-better criteria) into outputs (i.e. more-is-better criteria) in the most efficient way. Thanks to its rather generic nature, the concept of efficiently identifying DMUs turned out to be perfectly applicable within the domain of diet modelling, where DMUs become individual diets that can be evaluated based on the intake of multiple less-is-better (unhealthy), and more-is-better (healthy) nutrients (Kanellopoulos et al., 2018).

DEA aims to compare each diet with all other diets in a sample, and identify those that are efficient i.e. those that for certain levels of less-is-better nutrients contain the highest levels of more-is-better nutrients (i.e. output-oriented DEA model; OO-DEA), or those that for certain levels of more-is-better nutrients contain the lowest levels of less-is-better nutrients (i.e. input-oriented DEA model; IO-DEA) (Kanellopoulos et al., 2018). By doing so, it effectively builds an "efficiency frontier" (a convex hull) around inefficient diets, populated by the efficient diets. The efficient diet are considered to be the healthiest diets in the sample. The efficiency score (ES; $ES \in [0, 1]$) expresses how far away a certain inefficient diet is from the frontier. Every inefficient diet has $ES < 1$, while every efficient diets has $ES = 1$. These scores are obtained through linear programming (LP), the principal workhorse behind DEA method (Cooper et al., 2004; Kanellopoulos et al., 2018).

Given that diets consist of nutrient intake values that can be represented as real numbers, DEA operates in the vector space over the field of real numbers (i.e. a real vector space), which by definition supports two operations, namely addition and scalar multiplication, which further implies that every single inefficient diet can be projected onto the frontier. The projection can be done in many ways, depending on the choice of the distance function. Furthermore, it should be noted that every single vector that dominates a particular inefficient diet, presents a healthier alternative, which might actually be more preferable than the healthiest alternative. This is precisely what the

authors in (Kanellopoulos et al., 2018) explored - instead of constraining the set of solutions to the subset of Pareto-efficient diets that dominate a certain inefficient diet, they relaxed this constraint to include all diets that dominate it, and then searched for a diet within that subset that minimizes the absolute deviation between their food item intake values (the absolute loss function).

However, modeling preferability solely through minimization of the absolute loss leaves a lot to be desired. For instance, raw data might not serve as the most suitable representation, thereby making the extraction of useful information harder than necessary, or even impossible (Bengio et al., 2013). Furthermore, the absolute loss will always increase as much as possible in case both diets do not contain the same food item, while completely disregarding the fact that i. a particular food item (F_a) might serve as a substitute for another food item that is present in the diet that does not contain F_a, i.e. disregards food item similarities, or ii. that the consumer whose diet does not contain this food item, might actually prefer it (with a certain probability), which could be inferred from other food items that his diet does contain. In the section 4 we propose one possible alternative methodology that would arguably successfully address the drawbacks of the current method.

Furthermore, it seems reasonable to impose diet healthiness as the primary goal of diet optimization, as in the case above, however after making sure that our algorithm cannot produce a solution that is less healthy than the current one, we can impose additional search criteria, thereby sacrificing improvements in diet healthiness to a certain degree, for the purpose of maximizing some other criteria. Naturally, the criteria can be weighted in a way that would reflect the modeler's belief about their importance. Besides the preferability metric, some form of a sustainability metric would be arguably beneficial. The reasons for that, and how to include it in the DEA model will be explained in the following section.

3 Sustainability metric

Because of the exponential population growth, it is estimated that the global food production needs to increase at a rate of 1.2% per year to realize adequate diets for all people (Tilman et al., 2011). It is hard to imagine how much of an impact would that increase have on our ecosystem, given that current diets and production practices are already degrading terrestrial and aquatic ecosystems, depleting water resources, and driving climate change. More specifically, current food production and consumption patterns in Europe are held responsible for more than 25% of greenhouse gas emissions (GHGE), and more than 80% of arable land globally (Godfray et al., 2010). Shifting towards more sustainable diets is one of the strategies that could potentially remedy this issue. However, to be able to assess diet sustainability, there has to be some form of a sustainability metric assigned to every food item. Indeed, that poses a significant challenge. Nevertheless, many have tried, albeit with limited success. Delving deeper into this topic is well beyond the scope of this report, therefore the reader is referred to (Mertens et al., 2017).

Within the context of the SUSFANS project, greenhouse gas emissions (GHGE) and land use have been selected as relevant sustainability metrics for which reasonable amount of reliable data is available. Values for these metrics have been assigned to individual food items consumed in Europe as they occur in the FOODEX2 classification system by the European Food and Safety Authority (EFSA) (Mertens et al., 2018)

From the modelling standpoint, the sustainability indicator will change DEA's computational procedure in the same way as the aforementioned preferability criterion. More precisely, DEA will again shift its priority from finding the healthiest alternative, to finding the most sustainable one, that is at least as healthy as the original diet. Naturally, there can not be multiple criteria, all with maximal importance, so in case of having more than one additional criterion, a modeller will

have to impose a weighting scheme over the criteria. This weighting scheme should reflect the modeller's belief about their relative importance. To give a very simple illustrative example, in case of having two search criteria (e.g. preferability (P) and sustainability (S)) the modeller can induce a (fairly trivial) probability distribution over them, such as the one below:

$$\beta \cdot P + (1 - \beta) \cdot S$$

where β represents the probability mass, i.e. the importance weight.

4 Diet preferability

As mentioned before, DEA diet model developed in (Kanellopoulos et al., 2018) incorporates the preferability dimension in two way: by assigning consumers into groups (according to their age/sex) that are being optimized independently, and by applying the absolute loss similarity metric during the benchmarking procedure. As stated in the section 2, this section will propose the computational procedure for inferring consumer preferences, which is mainly addressing the potential issues of the absolute loss metric, and how to incorporate its results into the core DEA diet model.

There are at least two questions that one can pose while trying to define food item preferability, such as i. why consumers like the food items they buy?, and ii. which food items might consumers like? The first questions is complex, and giving a reasonably precise answer is highly unlikely. However, if observed on a macro level, human decision-making process could possibly be viewed as a Hidden Markov model (Baum & Petrie, 1966), i.e. a decision-making model that cannot be analyzed directly, but only through decisions that it makes. Furthermore, if those decisions are recorded over time, they could serve as a basis for predicting the answer to the second question. Moreover, if one augments that data with data from other sources similar (or complementary) to the original one, the prediction power should further increase. This idea has been used for many years, especially in the context of recommender systems, whose task is to turn data on users and their preferences into predictions of users' possible future likes and interests (Lü et al., 2012). Such models have been very successfull, as witnessed by social and technological impact of companies that leverage them, namely Google (Covington et al., 2016), Facebook (Ching et al., 2015), and Amazon (Linden et al., 2003).

More precisely, the problem of inferring consumer preferences can be cast as a collaborative filtering problem. Collaborative filtering (CF) is a class of methods for predicting a user's preference or rating of an item, based on his/her previous preferences or ratings and decisions made by similar users (Aggarwal, 2016; Sarwar et al., 2001).

There exist many types of collaborative filtering/recommender systems, which could be roughly categorized into 5 groups: neighborhood-based and model-based collaborative filtering, and content-based, knowledge-based, and ensemble-based/hybrid recommender systems (Aggarwal, 2016). That being said, the set of feasible methods will strongly depend on the amount and type of data, and the available computing power.

Given the data set used in (Kanellopoulos et al., 2018), which could be considered extremely small and sparse by today's standards, we decided to apply a neighborhood-based collaborative filtering approach, which builds on the premise that similar users (consumers) display similar patterns of behavior (e.g. purchase behavior) (Aggarwal, 2016). Our proposed algorithm consists of two computational layers: a data preprocessing layer ("PrepLayer"), and a similarity computation layer ("SimLayer"). The PrepLayer provides us with the ability to extract and organize the discriminative information from the data, i.e. to identify and disentangle the underlying explanatory

factors hidden in the observed (raw) data (Bengio et al., 2013). The SimLayer computes similarities between all pairs of food items, effectively building a "food item graph", which can be leveraged in many different ways.

Before presenting the details of our proposed algorithm, we will first explain the data set that the modelling had been performed on.

5 Data & Methods

5.1 Data

The data set that we used is the so-called "Nutrition Questionnaires plus" data set (NQplus) (Feskens, 2016), which includes the assessment of habitual dietary intake of 1784 participants (gathered through food frequency questionnaires and repeated 24-hour recalls), among other things. Essentially, the data set could be represented as a $n \times p \times d$ tensor T where $n = 1784$ is the number of participants/consumers, $p = 1297$ is the number of distinct food items, while $d = 3$ is the number of randomly sampled days, which were not necessarily the same for every consumer. Furthermore, every element of T is the amount (in grams) of a food item p_i, consumed by a consumer n_i on a day d_i.

For our modelling purposes, we decided to transform the original tensor T into the $n \times p$ matrix M, by averaging the $n \times p$ entries across d days. Stated differently, every entry of the matrix M can be thought of as the expected value of a consumer n_i's intake of a food item p_i.

5.2 Data preprocessing layer (PrepLayer)

As stated before, a suitable data representation (i.e. features) can greatly facilitate the extraction of useful information. The representation can be either handcrafted, or learned from data. Unfortunately, our data set (Feskens, 2016) does not allow for automatic representation learning, mainly because of its insufficient size. Therefore, we had to resort to handcrafting the representation.

One can think of this procedure as a way of increasing the signal-to-noise ratio. The goal was to obtain a weight matrix that would transform the raw data in such a way that every entry of M becomes a better approximation of the interrelations between food items and consumers.

Our approach has been inspired by the methods used in the scientific fields of natural-language processing and text mining, namely the ones for modelling semantic similarity between documents, which poses a significant theoretical challenge (Pincombe, 2004). If we think of consumers as documents, and food item quantities (or some transformation thereof) as word occurrence probabilities, these two seemingly completely different research fields become much more similar.

A variety of different approaches have been developed for modeling text document similarity (Pincombe, 2004), while most of them are essentially combinations of local and global weighting functions. In the context of diet modelling, a local weight should reflects the specificity of a certain food item within a certain consumer's diet, while a global weight should reflect the specificity of a certain food item across all consumers' diets. Furthermore, it has been noted that altering the local weighting function makes relatively little difference, but that changing the global weighting function does make a significant difference (Pincombe, 2004). However, it should be noted that the success of such functions can only be assessed in terms of their ability to emulate human judgments of importance. Therefore, objectively assessing the quality of the results poses a challenge.

We adopted the basic concept of having a combination of local and global weighting functions, however instead of selecting one of the commonly utilized global weighting functions that either do not have provably desirable properties, or might not be easily tweaked so as to accommodate

different data type, we decided to propose a new weighting function, which is (very) loosely based on the concept from theoretical artificial intelligence, namely the algorithmic/Solomonoff/universal probability distribution (Solomonoff, 1964), which is used for assessing the universal intelligence of an agent, and does have provably desirable properties (Legg, 2008). A thorough treatment of the entire subject is well beyond the scope of this technical report, therefore the interested reader is referred to (Solomonoff, 1964; Legg, 2008; Hutter, 2005).

In a nutshell, an agent's success within an environment, e.g. a human player's success in a video game, is weighted by the complexity of that environment. Stated differently, the success in a more/less complex environment carries more/less weight. This weighting scheme is generally considered to be a formalization of the philosophical principle called the "Occam's razor" (Rasmussen & Ghahramani, 2001), and could be formally expressed in the following way (for a single environment):

$$V_{\mu}^{\pi} \cdot 2^{-K(\mu)}$$

where V_{μ}^{π} represent the success of the agent π in the environment μ, while $K(\mu)$ stands for the Kolmogorov complexity of the environment μ. The expression $2^{-K(\mu)}$ represents the Occam's razor. As with the universal distribution, the concept of the Kolmogorov complexity is also beyond the scope of this report, therefore the reader is referrred to (Legg, 2008).

In the context of diet modelling, we can think of V_{μ}^{π} as the intake of the food item μ by the consumer π, e.g. how much probability mass is concentrated on this food item (i.e. local weight), while $2^{-K(\mu)}$ could represent the "complexity" of the food item μ, i.e. how much information on average does it carry, with respect to all other consumers. We will demonstrate our weighting function on a toy 4×5 matrix, however the concept generalizes to any other matrix, including our consumer-food matrix M.

We start with the raw matrix M_{toy}, such as the one below

$$\begin{bmatrix} 150 & 60 & 34 & 0 & 0 \\ 120 & 20 & 12 & 12 & 10 \\ 114 & 0 & 87 & 0 & 1 \\ 117 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where every row i represents a single consumer, every columns j represents a single food item, and every element M_{ij} expresses the amount of food item j (in grams) consumed by the consumer i. Given that different foods tend to have different scales (e.g. the average intake of oregano is much lower than the average intake of bread), it is advisable to normalize the columns. We apply the so-called "max abs scaling", i.e. divide every column by its maximum absolute value, so as to set the column range to $[0, 1]$ (given that we do not have negative values). The resulting matrix M_{MaxAbs} looks like this:

$$\begin{bmatrix} 1 & 1 & 0.3908 & 0 & 0 \\ 0.8 & 0.3333 & 0.1379 & 1 & 1 \\ 0.76 & 0 & 1 & 0 & 0.1 \\ 0.78 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Next, we compute two matrices, by transforming the rows/columns of M_{MaxAbs} into probabilities, i.e. dividing each row/column by its sum

$$\begin{bmatrix} 0.4183 & 0.4183 & 0.1635 & 0 & 0 \\ 0.2446 & 0.1019 & 0.0422 & 0.3057 & 0.3057 \\ 0.4086 & 0 & 0.5376 & 0 & 0.0538 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0.2994 & 0.75 & 0.2556 & 0 & 0 \\ 0.2395 & 0.25 & 0.0902 & 1 & 0.9091 \\ 0.2275 & 0 & 0.6541 & 0 & 0.0909 \\ 0.2335 & 0 & 0 & 0 & 0 \end{bmatrix}$$

ending up with M_{Prows} (left) and M_{Pcols} (right). M_{Prows} will serve as the local weight. Next, we start computing the global weight matrix, which we will refer to as the Occam's razor matrix, M_{Razor}. First, we compute the average information content of every column in the matrix M_{Pcols}. Shannon's information theory provides a suitable formalism for quantifying the average information content. The Shannon information content (h) is formally defined as

$$h(X = x_i) \equiv log_2 \frac{1}{p_i}$$

where $X = x_i$ is an event, while p_i is its probability. The *information entropy* (H) is a natural extension of the Shannon's information content (given the favourable properties of the logarithm function), which represents the average information content of a set of outcomes. Mathematically,

$$H(X) = \sum_i p_i log_2 \frac{1}{p_i}$$

where $X = x_i$ is an event, while p_i is its probability (Shannon & Weaver, 1963). In our case, the events are the food items, while the elements of M_{Pcols} represent their probabilities across the consumers. After computing the entropies of all columns, we end up with a single row, which we duplicate n times, where n is the number of rows of our original matrix M_{toy}. We get the matrix below:

$$\begin{bmatrix} 1.9908 & 0.8113 & 1.2167 & 0 & 0.4395 \\ 1.9908 & 0.8113 & 1.2167 & 0 & 0.4395 \\ 1.9908 & 0.8113 & 1.2167 & 0 & 0.4395 \\ 1.9908 & 0.8113 & 1.2167 & 0 & 0.4395 \end{bmatrix}$$

This will represent the K in $2^{-K(\mu)}$, i.e. the "environment complexity". It should be noted that we are slightly abusing the notion of complexity, in the sense that we treat the most uniform distribution as the least complex (as if it was just a string of numbers), while in reality it is actually the most complex. However, the way in which it is computed serves our purpose. That is, if the intake of a certain food items is more-or-less equally distributed among the consumers, the entropy will be high, which will in turn significantly decrease the value of $2^{-K(\mu)}$. Conversely, if the distribution is highly non-uniform, the entropy will be low, thereby increasing the value of $2^{-K(\mu)}$. If the entropy is equal to zero (i.e. if the food item has been consumed by only a single consumer), $2^{-K(\mu)}$ will be equal to 1, therefore there will be no penalty. We argue that this global weighting function correlates with what humans would perceive as important.

$$\begin{bmatrix} 2^{-1.9908} & 2^{-0.8113} & 2^{-1.2167} & 2^{0} & 2^{-0.4395} \\ 2^{-1.9908} & 2^{-0.8113} & 2^{-1.2167} & 2^{0} & 2^{-0.4395} \\ 2^{-1.9908} & 2^{-0.8113} & 2^{-1.2167} & 2^{0} & 2^{-0.4395} \\ 2^{-1.9908} & 2^{-0.8113} & 2^{-1.2167} & 2^{0} & 2^{-0.4395} \end{bmatrix} \rightsquigarrow$$

$$\begin{bmatrix} 0.2516 & 0.5698 & 0.43026 & 1 & 0.7373 \\ 0.2516 & 0.5698 & 0.43026 & 1 & 0.7373 \\ 0.2516 & 0.5698 & 0.43026 & 1 & 0.7373 \\ 0.2516 & 0.5698 & 0.43026 & 1 & 0.7373 \end{bmatrix}$$

Now that we've obtained our M_{Razor} matrix, the only thing that's left is to multiply it with M_{Prows} to obtain the weighted representation of our data (M_W), whose rows can further be converted into probabilities, so as to facilitate the comparison between the weighted representation, and the original probabilities M_{Prows}. M_{Prows} (top left), M_{Pcols} (top right) and M_{Wprobs} (bottom) are presented below:

$$\begin{bmatrix} 0.4183 & 0.4183 & 0.1635 & 0 & 0 \\ 0.2446 & 0.1019 & 0.0422 & 0.3057 & 0.3057 \\ 0.4086 & 0 & 0.5376 & 0 & 0.0538 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0.2994 & 0.75 & 0.2556 & 0 & 0 \\ 0.2395 & 0.25 & 0.0902 & 1 & 0.9091 \\ 0.2275 & 0 & 0.6541 & 0 & 0.0909 \\ 0.2335 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.2542 & 0.5758 & 0.1699 & 0 & 0 \\ 0.0919 & 0.0868 & 0.0271 & 0.4570 & 0.33701 \\ 0.2750 & 0 & 0.6188 & 0 & 0.106 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For instance, if we observe the elements $M_{Prows}(1,1)$ and $M_{Prows}(1,2)$, it seems that the consumer 1 equally prefers the food items 1 and 2, however according to those food items' columns $M_{Pcols}(1,:)$ and $M_{Pcols}(2,:)$, the preference of the food item 1 should not be that significant, while the preference of the food item 2 might actually be very significant. This is reflected in the M_{Wprobs} matrix, in which the value of food item 1 dropped from 0.4183 to 0.2542, while the value of food item 2 increased from 0.4183 to 0.5758. We can also note the interesting changes that took place at the locations $M_{Prows}(2,4)$ and $M_{Prows}(2,5)$, etc.

5.3 Similarity computation layer (SimLayer)

In this subsection we are going to explain the second computational layer of our proposed algorithm, namely the similarity computation layer. Its main function is to compute pairwise food item similarities, which would ultimately result in a "food item graph" - a fully connected graph that can be queried in many ways. Our primary intention was to explain the way in which such a graph could support the prediction of food item ratings by consumers that have not reported consuming those food items, which would allow for much more diverse recommendations, and potentially more accurate preferability estimation.

Because of our strong belief that the data set is fairly inaccurate due to the extremely low sampling frequency (i.e. 3 days/year), we have decided to binarize our data set by applying the thresholding procedure, which replaces each value in our weighted representation matrix $M_{W}probs$ with 0 if the value is less than some fixed constant T, or 1 if the value is greater than that constant. By doing so, we essentially stated that although the values of our data could tell us something, it might be a good thing to develop a layer that is fairly robust, and not fully dependent on these observed, and likely noisy values. For the purpose of this report, the binarization was performed over the weighted representation matrix M_{Wprobs}, while the threshold was set to 0. Given our M_{toy} matrix, the resulting matrix would look like the one below:

$$\begin{bmatrix} 0.2542 & 0.5758 & 0.1699 & 0 & 0 \\ 0.0919 & 0.0868 & 0.0271 & 0.4570 & 0.33701 \\ 0.2750 & 0 & 0.6188 & 0 & 0.106 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Naturally, the parameter T can be tuned. Furthermore, it is possible to introduce a parameter N_{top}, that would transform only the top N values of each row of the weighted representation matrix M_{Wprobs} into 1's , while the rest would be 0's.

The similarity metric that we selected is based on the information-theoretic concept of mutual information (MI or $I(X;Y)$). The motivation for considering mutual information is its capability to measure a general dependence among random variables (i.e. food items)(Zhou et al., 2004), as opposed to some other commonly employed correlation metrics (e.g. Pearson correlation coeff.), which measure only linear dependence.

Intuitively, the mutual information $I(X;Y)$ between two random variables X and Y is the amount of information Y (X) gives about X (Y), i.e. how much knowing one of these variables reduces uncertainty about the other. Mathematically, we define this as the difference between the entropy of X and the entropy of X conditioned on Y.

$$I(X;Y) = H(X) - H(X|Y)$$

The mutual information is symmetric, i.e. $I(X;Y) = I(Y;X)$, as shown below:

$$H(X) + H(Y|X) = H(X,Y) = H(Y,X) = H(Y) + H(X|Y)$$

$$H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$I(X;Y) = I(Y;X)$$

Do note that in this case, the notion of entropy is used properly. Furthermore, $I(X;Y)$ can be expressed via the Kullback-Leibler divergence, a fundamental equation of information theory that quantifies the proximity of two probability distributions. Essentially it quantifies in bits how close a probability distribution P is to a model (or candidate) distribution Q. Mathematically,

$$D_{KL}(P\|Q) = \sum_i P(i)\ log_2 \frac{P(i)}{Q(i)}$$

In the case of mutual information, P represents the joint probability distribution $p(x,y)$, while Q represents the product of their marginal distributions, $p(x)p(y)$:

$$I(X;Y) = D_{KL}(p(x,y)\|p(x)p(y))$$

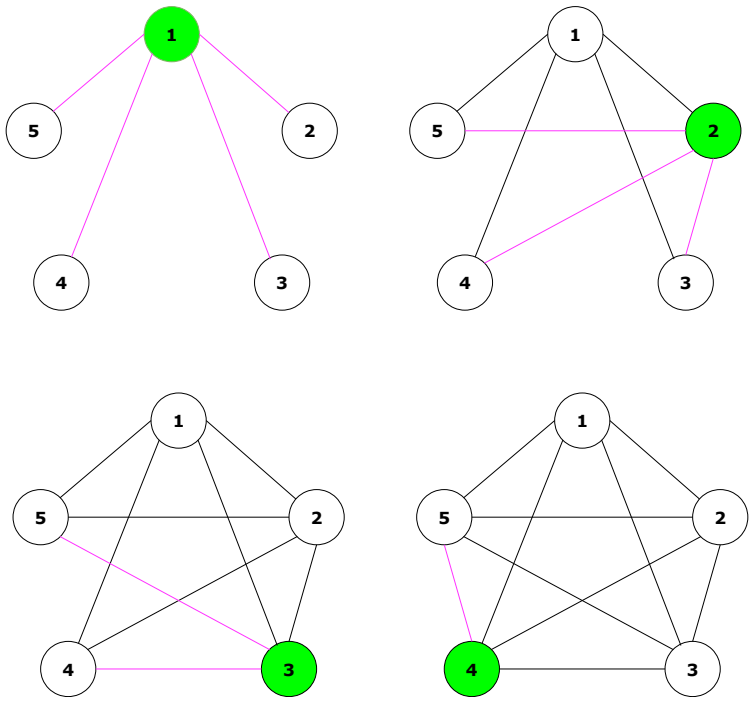
which arguably justifies its selection as a dependence metric.

Furthermore, it should be noted that the value of $I(X;Y)$ can be very small, even if X and Y are highly correlated, since their entropies can be small, therefore we normalize it by the maximum value between $H(X)$ and $H(Y)$:

$$\bar{I}(X;Y) = \frac{I(X;Y)}{max[H(X), H(Y)]}$$

The SimLayer's computational procedure is fairly straightforward, in the sense that the algorithm takes in a single food item vector as input, and computes its similarity with all other food item vectors. This procedure is performed $p(p-1)/2$ times, where p is the number of distinct food items. Essentially, all that it needs to do, is to compute the values of triangular entries of an $p \times p$ matrix, namely because the similarity metric is symmetric, and the values of the diagonal will always be equal to 1. i.e. a variable X shares 100% of its information content with itself.

For instance, the similarity computation between the food item (column) vectors of our binarized M_{Wprobs} matrix, and the resulting food item similarity graph are shown below:



As with the PrepLayer's output, the assessment of the SimLayer's output quality is challenging, mainly because the choice of the similarity metric is to some extent subjective, and its performance can be evaluated either through direct manual inspection, or indirectly, i.e. if its utilization increases the value of some other variable that we want to maximize, e.g. when better recommendations result in higher revenue etc. Indeed, our goal is to maximize diet preferability, however given that our model does not have a feedback loop (i.e. there is no feedback from the consumers that receive our model's recommendations), we had to resort to manual verification of the results.

We selected a few food items from our data set, for which we believed it would be easy to verify the quality of the results. The results are ranked according to their similarity with the target food item, in descending order. The reader has to keep in mind that these food item similarities are not based on food features (e.g. nutrient content, taste, texture, etc.), but on consumers that bought them:

- *Soy drink - natural*
  - Soy drink (original), linseed, soy drink (various flavors), eggplant (cooked), vegetable yoghurt (fruits/vanilla), apricots (dried), vegetarian hamburger

- *Coconut bread*
  - Wheat bran, hazelnuts (unsalted), muesli (fruits), meat with salad (lunch), raisins (dried), cracked wheat bran
- *Coffee milk (half-and-half)*
  - Coffee milk (whole), coffee milk (vegetable, half-and-half), champignons, sugar, vla vanilla (whole), breakfast product (Brinta), coffee creamer, milk (half-and-half)
- *Dried oregano*
  - paprika (powdered), tomato ketchup, tomatoes (mashed), red paprika (cooked), pepper (black/white), stock cube, onions (cooked)

Furthermore, it is also interesting to observe the differences between e.g. food items consumed by the majority of the consumers within the sample, and the ones that are not. For instance, white bread has been recorded in approx. 22.3% of the consumers' diets, whereas whole grain bread with nuts has been recorded only in approx. 0.78% of the consumers' diets. Some of their most similar food items are the following:

- *White bread*
  - Margarine 80% fat, ham (shoulder), whole grain bread, white bread (milk), Gouda cheese, vegetables (raw), tiger bread, Brie cheese, broccoli (cooked), chicken eggs
- *Whole grain bread with nuts*
  - Currant bread (wholegrain, morels, wholegrain bread, lychee, cereal bar, fennel (raw), oat cake, vegetarian mince (unprepared), cress sprouts (raw)

The food item graph produced by the SimLayer can be queried in many ways. For instance, it is possible to search for a food item that is the most similar to a set of food items, or to find maximal weighted cliques (i.e. sets in which all food items are strongly correlated with each other), or to cluster food items by minimizing mutual information between clusters, etc.

The final computational step that our algorithm performs is rating prediction. That is, it predict the probability that a certain consumer will prefer a particular food item that has not been observed in his data. Formally, the prediction is computed in the following way:

$$r_{ci} = \frac{\sum_{j \in N} s_{ij} r_{cj}}{\sum_{j \in N} s_{ij}}$$

where r_{ci} is the rating of the food item i by the consumer c, s_{ij} is the similarity of the food items i and j, r_{cj} is the rating of the food item j by the consumer x (i.e. the values of our M_{Wprobs} matrix), while N is the subset of food items that have been consumed (rated) by the consumer x. The subset can contain all food items consumed by a particular consumer (i.e. a full set), or the top n food items that are the most similar to the food item i.

References

Aggarwal, C. C. (2016). *Recommender systems: The textbook* (1st ed.). Springer Publishing Company, Incorporated.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *Ann. Math. Statist.*, *37*(6), 1554-1563.

Bengio, Y., Courville, A., & Vincent, P. (2013, August). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, *35*(8), 1798–1828. doi: 10.1109/TPAMI.2013.50

Ching, A., Edunov, S., Kabiljo, M., Logothetis, D., & Muthukrishnan, S. (2015, August). One trillion edges: Graph processing at facebook-scale. *Proc. VLDB Endow.*, *8*(12), 1804–1815. doi: 10.14778/2824032.2824077

Cooper, W. W., Seiford, L. M., & Zhu, J. (2004). Data envelopment analysis. In *Handbook on data envelopment analysis* (pp. 1–39). Springer.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th acm conference on recommender systems.* New York, NY, USA.

EFSA. (2017, December). Dietary Reference Values for nutrients Summary report. *EFSA Supporting Publications*, *14*(12), n/a–n/a. doi: 10.2903/sp.efsa.2017.e15121

Feskens, G. A., E. J. M. (2016). Nqplus. *Wageningen University Research, Afdeling Humane Voeding.*

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., … Toulmin, C. (2010, February). Food Security: The Challenge of Feeding 9 Billion People. *Science*, *327*(5967), 812–818. doi: 10.1126/science.1185383

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). *The elements of statistical learning : data mining, inference, and prediction.* New York: Springer. (Autres impressions : 2011 (corr.), 2013 (7e corr.))

Hutter, M. (2005). *Universal artificial intelligence: Sequential decisions based on algorithmic probability.* Springer.

Kanellopoulos, A., Ivancic, A., Geleijnse, M., Bloemhof-Ruwaard, J., & Kuijnsten, A. (2018, May). *Deliverable 7.3: Initial model for designing SHARP diets | SUSFANS.*

Legg, S. (2008). *Machine super intelligence.* Lugano: University of Lugano.

Linden, G., Smith, B., & York, J. (2003, January). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80. doi: 10.1109/MIC.2003.1167344

Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012, October). Recommender Systems. *Physics Reports*, *519*(1). (arXiv: 1202.1112) doi: 10.1016/j.physrep.2012.02.006

Macdiarmid, J. I., Kyle, J., Horgan, G. W., Loe, J., Fyfe, C., Johnstone, A., & McNeill, G. (2012, September). Sustainable diets for the future: Can we contribute to reducing greenhouse gas emissions by eating a healthy diet? *The American Journal of Clinical Nutrition*, *96*(3), 632–639. doi: 10.3945/ajcn.112.038729

Mertens, E., Kuijsten, A., Zanten, H. V., Kaptijn, G., Dofková, M., Mistura, L., … Geleijnse, J. (2018). Diet composition and environmental impact in four european countries. *Paper in preparation, SUSFANS deliverable 7.2.*

Mertens, E., Veer, P. v., Hiddink, G. J., Steijns, J. M., & Kuijsten, A. (2017, March). Operationalising the health aspects of sustainable diets: a review. *Public Health Nutrition*, *20*(4), 739–757. doi: 10.1017/S1368980016002664

Pincombe, B. (2004, September). *Comparison of human and LSA judgements of pairwise document similarities for a news corpus* [Public Announce].

Rasmussen, C. E., & Ghahramani, Z. (2001). Occam's razor. In *In advances in neural information processing systems 13* (pp. 294–300). MIT Press.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on world wide web* (pp. 285–295). New York, NY, USA: ACM. doi: 10.1145/371920.372071

Shannon, C. E., & Weaver, W. (1963). *A mathematical theory of communication*. Champaign, IL, USA: University of Illinois Press.

Solomonoff, R. J. (1964). A formal theory of inductive inference. part i. *Information and control*, *7*(1), 1–22.

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011, December). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), 20260–20264. doi: 10.1073/pnas.1116437108

Zhou, X., Wang, X., Dougherty, E. R., Russ, D. E., & Suh, E. (2004). Gene clustering based on clusterwide mutual information. *Journal of Computational Biology*, *11*(1), 147–161. doi: 10.1089/106652704773416939