Rubrique

Towards a recommender system for healthy nutrition

An automatic planning-based approach

Ngoko Yanik

Laboratoire d'Informatique de Paris Nord
University of Paris 13
Paris
France
yanik.ngoko@lipn.univ-paris13.fr

RÉSUMÉ. Ce papier introduit le problème de la planification automatique des repas. Étant donné un ensemble de repas caractérisés par une composition calorifique, un prix, des éléments de qualité nutritionnelle et la période de consommation (petit-déjeuner, dîner etc.), le but est de construire l'affectation optimale des repas dans les périodes afin d'optimiser la qualité de l'alimentation et/ou le plan. Le problème de la planification automatique des repas est particulièrement intéressant en Afrique. Avec la récente émergence des systèmes de télécommunications ainsi que la construction des bases de données sur l'alimentation africaine, il peut servir à construire des systèmes de recommandation, permettant aux consommateurs d'optimiser leur budget de nutrition tout en maintenant une alimentation équilibrée. Ce papier contribue sur ce défi en formalisation le problème, l'analysant et en proposant un algorithme par séparation-évaluation pour sa résolution. Enfin, nous effectuons une validation expérimentale à partir des données sur l'alimentation en Tanzanie. Les résultats montrent que nous pouvons produire des plans nettement meilleurs que ceux issus d'approches plus naïves.

ABSTRACT. In this paper, we focus on the automatic meal planning problem (AMP). Assuming a set of meals characterized by a calorific composition, a price, a nutrient composition and mealtimes, the objective is to decide on the meal to assign to each mealtime such as to obtain an optimal plan in term of nutritional quality and prices. AMP is particularly interesting in Africa. Indeed, thanks to the emergence of telecommunication networks and works done on the statistical modeling of nutrition in Africa, it can serve for designing recommender systems, accessible on cellular phones that will help consumers to make a better planning of their budget while keeping a balanced nutrition. Our study contributes to this objective by formalizing the problem, analyzing it and providing a branch and bound algorithm for its resolution. Finally, we did an experimental evaluation based on open data available for nutrition in Tanzania. The results show that the plan we produce can largely overpass naive solutions.

MOTS-CLÉS : Système de recommandation, alimentation équilibrée, problème d'optimisation de contraintes

KEYWORDS : Recommender system, balanced nutrition, constraint optimization problem

1. Introduction

The global objective of this study is to help to improve the world nutrition by providing adequate numeric tools for a balanced and responsible nutrition. For this purpose, we introduce the automatic meal planning problem (AMP). Assuming a set of meals that we mainly characterize by their ingredients, nutrient composition, price and a window of mealtimes, the objective is to automatically build a meal plan that states *what to eat in each mealtime* while optimizing a budget and a nutritional balance.

The AMP is particularly interesting in African countries. Indeed, it can serve for building recommender systems that will help people to keep a balanced nutrition. The idea of building such recommender systems is not new. It is part of a general trend that consists of investigating nutrition challenges with computer algorithms. Some popular questions in this tendency are the finding of equivalence between ingredients [1], the analysis of flavor between recipes [2] or the discovery of the structure similarity in recipes [3]. In this paper, our focus is on the automatic composition of meal plans.

Closer to our objective, we can refer to the CHEF system [4], a recommender system based on user preference [5], the smart kitchen system [6] or the daily meal plan recommender system [8]. Our work shares several common features with these works. These are : the formalization of meals through cook recipes, the management of users preferences, the distinction of meals in types or the objective to converge towards a healthy nutrition. However our work differs from these studies on three main points. The first point is that these works tackle the problem mainly in an Artificial Intelligence (case-based reasoning, expert systems etc.), information system or data analysis perspective (data clustering, statistical analysis etc.) while we are interesting in modeling, formalizing and solving the combinatorial problem of meal plan composition. The second novelty that our work introduces is to consider qualitative evaluation of meals based on the a nutritional score system that we will refer to as the Hercberg score [9]. The Hercberg score is a classification that ranges foods in distinct classes depending on the quantity of nutrients they include and the type of ingredients they have. This classification is becoming a popular standard and was adopted for food labeling in France. Finally, we provided experiments that demonstrate how our system can be used for healthy nutrition based on open data available for nutrition in Tanzania.

The remainder of this paper is organized as follows. In Section 2 , we discuss the related work. Section 3 introduces the theoretical formulation of the automatic meal planning problem. In Section 4, we propose a heuristic for solving the problem and evaluate it. We conclude in Section 5.

2. Related work

The design of automatic meal planner was investigated rather early in Artificial Intelligence. One of the first proposed system was CHEF [4], a case-based reasoning system that was able to recommend dishes based on their types and the taste expected by the consumer. As the CHEF system, in the problem we propose, dishes are ranged in types and are characterized by their ingredients. Unlike CHEF, we do not explicitly account on the tasting but consider a general concept of preference. Finally, we consider the quality of meals based on the nutrient composition. In [5], the authors propose a recommender system for recipes based on users preferences. From several observations made on recipes chosen by the users, the system is able to detect what are the user favorite ingredients.

Based on these ingredients, a classification of recipes is proposed and then used for recommending recipes to users. In our work, we also handle user preferences. However, it is only a criteria for deciding on the best recipes. In [6] the authors introduce the smart kitchen, an intelligent kitchen that returns qualitative data about cooking processes. The smart kitchen includes sensors and camera that serve for detecting any cooking action and estimating its *nutritional and calorific value*. The system also provides recommendations for adjusting the real-time composition of a meal towards a nutritional balance. While the smart kitchen is a hardware and software innovation, we focus in this paper on the software aspect of meals planning. In [7] the authors propose a planning system for healthy nutrition. The system is based on propositional logic and can be used on mobile devices. As our work, the objective of this study is to propose a digital assistant to fill the lack of experts in poor countries. However, we differ from this study on our modeling of the meals planning problem. In [8] the authors propose a recommender system for building daily nutrition plan. They demonstrate that their solution can provide balanced nutrition plans that respect users' preferences. A common feature between our formulation and this recommender system is the idea of considering the quality of meal plans. But, while this paper proposes a custom classification of foods, we consider the Hercberg score. In addition, we do not only focus on daily plan (as it is the case in this study) : our mealtime window can include weeks and months. Finally, it is important to observe that our study is possible because of existing theoretical formulations for characterizing the quality of nutrition based on discrete quantity. These are for instance the Hercberg score [9] and nutrient composition tables [11]. In the next, we present our model.

3. Problem description and analysis

3.1. General view

We consider a family that has a finite set of mealtimes $D = \{t_1, t_2, \dots, t_k\}$. Typically, we might have $k = 3$ with t_1 being the breakfast, t_2 the lunch and t_3 the diner. We also assume that the family is interested in a meal plan in a horizon of $\Delta = \{1, \dots, T\}$ days. $T = 30, 1$ are meaningful values when considering cultural notion as the concept of "ration" in Africa [1]. At each mealtime, the family can opt for a meal issued from a finite set $M = \{m_1, m_2, \dots, m_n\}$. The general goal in AMP is to build an assignment σ that for each meal m_i, mealtime $t_u \in D$ and days $d_j \in \Delta$ is such that $\sigma(m_i, t_u, d_j) = 1$ if on day d_j and at the mealtime t_u, the meal m_i was chosen. The built assignment must satisfy objectives and criteria specified by the family. This general formulation is subject to constraints and objectives that we will define in the next.

3.2. Formal definition

For the sake of simplicity, we reduce the family to a single person. This choice will impact the formulation of constraints related to the nutritional balance. We also assume the following (additional) input data :

– K_i the calories provided by meal m_i ;

1. In several African countries, husbands give a budget for cooking to their wives everyday or at the beginning of the month.

– $\alpha_i^1, \alpha_i^2, \alpha_i^3$ the percentage of carbohydrates, fat and proteins in meal m_i ;

– p_i, the price of the m_i ;

– A Boolean function $\gamma(i, c)$ such that $\gamma(i, c) = 1$ is meal m_i belongs to culture c. We also assume that we have a set C of cultures.

– H, W, G, A the height (cm), weight (kg), gender, age of the person we consider ; $G = 1$ for female, 0 for male.

– $R = \{R_1, \ldots, R_5\}$ the classes of recommendations the user could follow. R_1 corresponds to a consumer that makes little or no exercise ; R_2 is a consumer that makes $1-3$ days of exercise per week ; R_3 a consumer with $3 - 5$ days of exercise ; R_4 a consumer with $6 - 7$ days of exercise ; and R_5 a consumer with very intensive exercises. We also assume the Boolean variables r_is that are such that $r_i = 1$ if the user chose the class R_i.

– $E(t_u)$ the set of acceptable meal in the mealtime t_u ;

– q_i , the Hercberg score of meal m_i ; the lower is q_i, the better is the quality of m_i.

We consider the percentage of proteins, fat and carbohydrates because as mentioned in [8], they are crucial for a balanced diet. We range each meal in a culture. This choice is among other things motivated by an observation made in prior studies [6] : consumers choose their dishes according to cultural preferences. With the set $E(t_u)$ of acceptable meals, our objective is to distinguish between types of meals that are appropriate depending on the mealtime. Finally, we consider the height and weight of the consumer because this will serve to estimate his requirement in term of calories. Assuming these data, in the next, we will now define the constraints.

3.2.1. Constraints

We consider the following constraints :

C_1 **: One meal per mealtime**

$$\forall d_j, t_u, \sum_{m_i \in M} \sigma(d_j, t_u, m_i) = 1$$

C_2 **: The meal must be accepted**

$$\forall d_j, t_u, \sum_{m_i \in M | m_i \notin E(t_u)} \sigma(d_j, t_u, m_i) = 0$$

C_3 **: Maximum budget limit per day**

$$\forall d_j, \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).p_i \leq B$$

(the daily budget for eating is B)

C_4 **: Diversity in meal choice**

$$(1) \forall m_i, \sum_{t_u \in D} \sum_{d_j \in \Delta} \sigma(d_j, t_u, m_i) \leq F_i; (2) \forall m_i, d_j \sum_{t_u \in D} \sigma(d_j, t_u, m_i) \leq 1$$

(1) means that a meal is chosen at most F_i times. (2) means that per day, a meal cannot be chosen twice.

C_5 **: Cultural preferences**

$$\forall c \in C, \sum_{d_j \in \Delta} \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).\gamma(i, c) \geq I_c;$$

(meal from a culture c will be chosen at least I_c times)

C_6 **: Calorific recommendation based on the Harris-Benedict equation**

$$\forall d_j, \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).K_i = [G.bmr_1 + (1-G).bmr_2].bmrFactor$$

Here, $bmr_1 = 447.593 + 9.247W + 3.098H - 4.330A$; $bmr_2 = 88.362 + 13.397W + 4.799H - 5.677A$ and $bmrFactor = 1.2r_1 + 1.375r_2 + 1.55r_3 + 1.725r_4 + 1.9r_5 + 200\epsilon_1$

This constraint expresses the calorific need according to the Harris-Benedict equation [10]. We added a margin error factor $\epsilon_1 \in [-1, 1]$ that ensures that the proposed plan will exceed or be lower of at most 200 calories from the standard recommendation.

C_7 **: Balanced diet requirement**

$$\forall d_j, \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).\alpha_i^1.K_i = (0.55 + 0.10\epsilon_2) \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i)K_i$$

$$\forall d_j, \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).\alpha_i^2.K_i = (0.275 + 0.75\epsilon_3) \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i)K_i$$

$$\forall d_j, \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i).\alpha_i^1.K_i = (0.225 + 0.125\epsilon_4) \sum_{t_u \in D} \sum_{m_i \in M} \sigma(d_j, t_u, m_i)K_i$$

Here, $\epsilon_2, \epsilon_3, \epsilon_4 \in [-1, 1]$. The idea in balanced diet requirement is to ensure that in the calories gained each day, between 45 to 65% come from carbohydrates, 20 to 35% from fat and 10 to 35% from proteins. Let us notice that these values are recommended by experts in nutrition [8].

3.3. Objective functions in the automatic meal planning problem

In AMP, we want to minimize the price and the Hercberg score of the plan : the lower is this score, the better is the quality. We modelize the price and quality of a plan as follows : $Price = \sum_{t_u \in D} \sum_{m_i \in M} \sum_{d_j \in \Delta} \sigma(d_j, t_u, m_i).p_i$; $Quality = \sum_{t_u \in D} \sum_{m_i \in M} \sum_{d_j \in \Delta} \sigma(d_j, t_u, m_i).q_i$. The objective function in AMP is the normalized function

$$Cost = \lambda \frac{Price}{|Price| + |Quality|} + (1-\lambda) \frac{Quality}{|Price| + |Quality|}$$

Here $\lambda \in [0, 1]$ is a parameter defined by the consumer to give more interest in either price or quality.

3.4. Analysis

It is straightforward to notice that AMP is a Constraint Optimization Problem. The interest in the observation is that we can therefore consider general Constraint optimization framework like Branch and Bound for its resolution. We also have the following result.

Theorem 3.1 *If we only consider the constraints C_1, C_2 and C_3 then AMP is NP-hard.*

The proof is given in the appendix. It is based on a reduction to the 3-partition problem. Finally, let us notice that several variants of AMP can be proposed. For instance, we can model the diversity in considering *neighbor meals*. A neighbor could refer to meals of the same day or those in consecutive days.

4. Heuristic and evaluation

From the mathematical formulation proposed in Section 3.2, we can derive an Integer Linear Program (ILP) for solving AMP. The only difficulty will could come from the nonlinear objective function. Despite the interest in ILP, let us notice that the runtime can quickly explode when we consider big problems. However, let us observe that we described AMP as a constraint optimization problem. For such problems, branch and bound algorithms (B&B) are efficient. We will describe such an algorithm in the next.

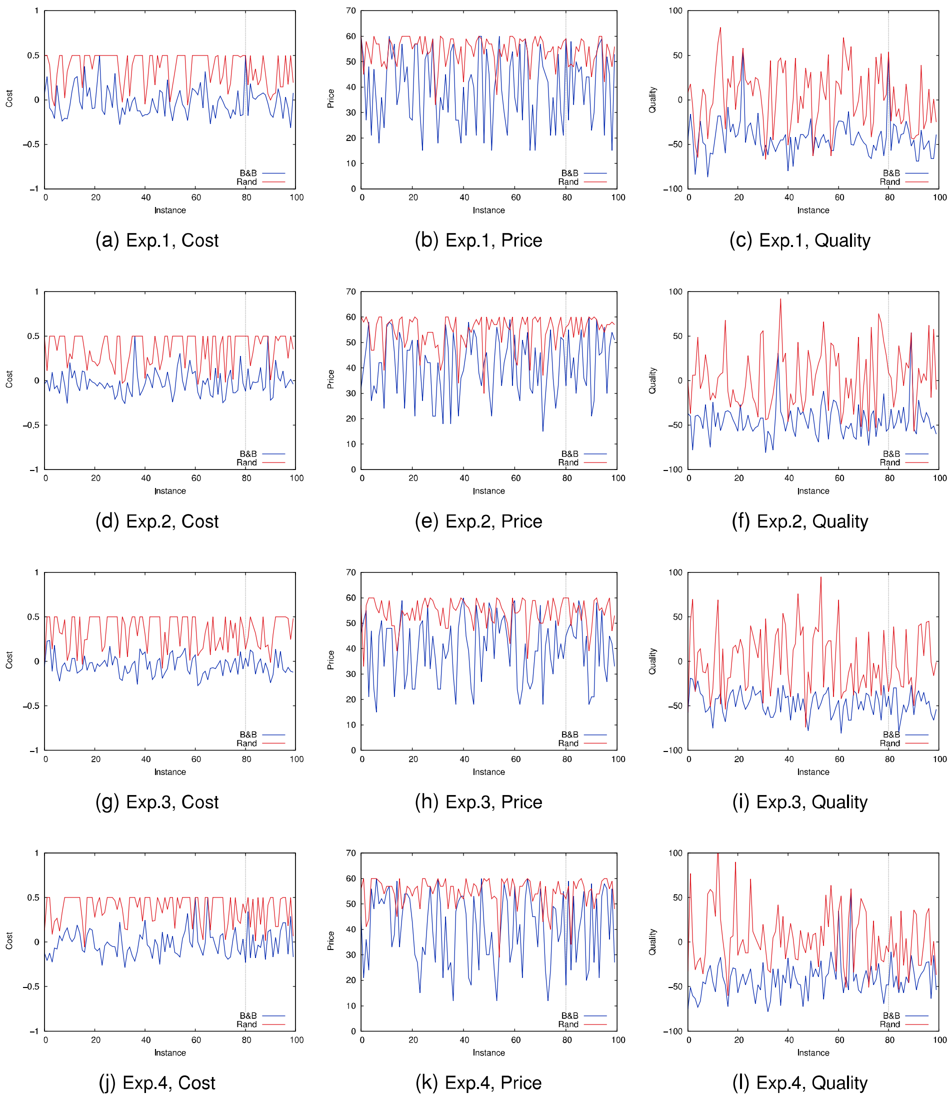
4.1. A branch and bound algorithm for AMP

In this algorithm, we consider that a solution to AMP is a one dimensional vector X such that each $X(e)$ states for a pair $e = (d_j, t_u)$ the meal m_i that was chosen. Consequently, $|X| = |D|.|\Delta|$ and the domain of possible values for $X(e)$ is $dom(X(e)) = E(t_u)$. In the B&B algorithm, we start by assigning a value to $X(0)$ and evaluate partially all the constraints from C_3 to C_7. For instance, the partial evaluation of C_3 consists of checking whether or not we already exceeded the maximal budget. If no violation is found, we continue in assigning a value to $X(1)$ and repeating the process. Let us now assume that at a moment, we have a sub-vector $X(1...i)$ and that we detect a violation with the assignment made to $X(i+1)$. Then, we backtrack by changing the value of $X(i+1)$. If no possible values could be assigned to $X(i+1)$ we backtrack to $X(i)$. Finally, in this algorithm we keep every time a lower bound : the partial value of $Cost$ for the assignment we made. If this bound exceeds the best found solution, we backtrack.

4.2. Experimental evaluation

We evaluated the B&B algorithm in using a database of Tanzanian food composition [11]. We chose from this database 106 recipes of Tanzanian meals for which we have the ingredients and nutrient composition. Based on these data, we computed the quality of each meal and their calorific values. In the experiments, we randomly generated the price of each meal in choosing a value between 1 and 50. We also assume that half of the recipes belong to one culture and the remaining to another one. We also fixed the following values $k = 3, T = 4, I_c = 3, F_i = 0.3.(3 \times 4), \lambda = 0.5$. Finally, we assumed different settings where the consumer has one of the *standard profile* defined in [12]. We chose 4 of these profiles : female sedentary, 31-50 (Exp. 1), female sedentary, 51+ (Exp.2), male sedentary, 51+ (Exp.3), male sedentary, 31-50 (Exp.4). For each experiment, we randomly generated 100 price distributions. We then compared the best solution obtained by the B&B algorithm after at most 5 min, with a randomized algorithm. This latter solution was obtained by running a randomized version of the B&B that was interrupted once a feasible solution was found. The randomization was applied here on the ordering we used for processing the $X(i)$s. The solution issued from the randomized algorithm could correspond to the consumer choice. Indeed, we do not believe that in practice, consumers will make a deep exploration of the huge space of potential solutions. Therefore, the first feasible solution (naive solution) could probably be the one they will adopt.

The results of our experiments are presented in Figure 1. As expected, the solutions of the B&B are better (in cost) than the naive ones. But more interestingly, they are not only better when considering the objective function : as showed by the curves on prices and quality, we are able to find plans that a both cheaper and of better quality. Let us recall indeed that in the Hercberg score, the lower is the score, the better is the quality.



(a) Exp.1, Cost (b) Exp.1, Price (c) Exp.1, Quality

(d) Exp.2, Cost (e) Exp.2, Price (f) Exp.2, Quality

(g) Exp.3, Cost (h) Exp.3, Price (i) Exp.3, Quality

(j) Exp.4, Cost (k) Exp.4, Price (l) Exp.4, Quality

Figure 1 – Cost, price and quality in different experiments

5. Conclusion

In this paper, we modeled the automatic design of balanced meals plans and proposed an algorithm for its construction. Our modeling is based on keys mathematical concepts in nutrition like the Harris-Benedict equation and the distribution of calories in healthy diet. We then validated our algorithm in considering a database of Tanzanian foods. The experimental results showed that with our modeling, we are able to find balanced nutrition plans that outperform naive solutions on both prices and quality. For continuing this work, we have three main perspectives. The first is to refine the modeling and evaluation in including other elements like tasting and enlarging the database of meals. The second is to validate the approach in considering a pool of real consumers. Finally, we envision to reduce the runtime of the B&B algorithm in using parallelism and advanced constraint optimization techniques.

6. Bibliographie

[1] YUKA SHIDOCHI, TOMOKAZU TAKAHASHI, ICHIRO IDE, HIROSHI MURASE, « Finding replaceable materials in cooking recipe texts considering characteristic cooking actions », *CEA '09 Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, pp. 9-14, 2009.

[2] YONG-YEOL AHN, SEBASTIAN E. AHNERT, JAMES P. BAGROW, ALBERT-LÁZLÓ BARABÁSI, « Flavor network and the principles of food pairing », *Nature, Scientific Reports*, vol. 1, n° 196, pp 1-7, 2011.

[3] LIPING WANG, QING LI, NA LI, GUOZHU DON, YU YANG, « Substructure Similarity Measurement in Chinese Recipes », *Proceedings of the International World Wide Web Conference*, pp. 979-988, 2008.

[4] KRISTIAN J. HAMMOND, « CHEF : A model of Case-based Planning », *Proceedings of AAAI*, pp. 267-271, 1986.

[5] MAYUMI UEDA, SYUNGO ASANUMA, YUSUKE MIYAWAKI, SHINSUKE NAKAJIMA, « Recipe Recommendation Method by Considering the User's Preference and Ingredient Quantity of Target Recipe », *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 519-523, 2014.

[6] JEN-HAO CHEN, PEGGY PEI-YU CHI, YUSUKE MIYAWAKI, HAO-HUA CHU, CHERYL CHIA-HUI CHEN, POLLY HUANG, « A Smart Kitchen for Nutrition-Aware Cooking », *Pervasive computing*, vol. 9, n° 4, pp. 58-65, 2010.

[7] FERNANDO ZACARIAS F., ROSALBA CUAPA, ERICK MADRID, DIONICIO ZACARIAS, « Healthy Nutrition Under ASP-PROLOG », *International Journal of Computer Networks & Communication*, vol. 5, n° 3, pp. 91-102, 2013.

[8] DAVID ELSWEILER, MORGAN HARVEY, « Towards Automatic Meal Plan Recommendations for Balanced Nutrition », *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 313-316, 2015.

[9] SERGE HERCBERG, « Propositions pour un nouvel élan de la politique nutritionnelle française de santé publique », *http ://www.sante.gouv.fr/IMG/pdf/rapport_Hercberg_15_11_2013.pdf*, Accessed 26 October 2015

[10] J. ARTHUR HARRIS, FRANCIS G. BENEDICT, « A Biometric Study of Human Basal Metabolism », *Published by The Carnegie Institute of Washington. Proc Natl Acad Sci U S A.* vol. 4, n° 12, pp 370-373, 1918.

[11] ZOHRA LUKMANJI, ELLEN HERTZMARK, NICOLAS MLINGI, VINCENT ASSEY, GODWIN NDOSSI, WAFAIE FAWZI, « Tanzania Food composition Tables », *http ://www.hsph.harvard.edu/nutritionsource/food-tables/* Accessed 26 October 2015.

[12] KATHLEEN M. ZELMAN, « Estimated Calorie Requirements », *http ://www.webmd.com/diet/estimated-calorie-requirement*, Accessed 26 October 2015

7. Appendix

7.1. Proof of theorem 3.1

Let us recall that in this proof, we consider a restricted version of AMP that only includes the constraints C_1, C_2, C_3. The NP-hardness proof is based on a reduction to the 3-partition problem. Given a set S of $3l$ positive integers $s_1, \ldots s_{3l}$, the objective in 3-partition is to subdivide S into l triplets $S_1, \ldots S_l$ such that the sum of number in each subset is equal and the sets $S_1, \ldots, S_l$ cover S.

From this instance, we propose to build the following AMP instance : We set $T = l$ and $k = 3$. This means that the AMP instance has 3 mealtimes per day and covers l days. We assume $3l$ meals and associate each meal m_i with the price $p_i = s_i$. We fix $F_j = 1$ (all chosen meals are distinct) and

$$B = \frac{\sum_{u=1}^{3l} e_i}{l}$$

Finally, we set the quality of each meal to 0 (such values exist in the Hercberg score).

For solving any instance of the 3-partition problem, we formulate the associated AMP instance and solve it. If σ is the solution, then we associate each S_j with a day d_j as follows :

$$S_j = \{p_i | \sigma(d_j, t_u, m_i) = 1\} (a)$$

It is straightforward to notice that if there is a solution to the 3-partition then there is a solution to the associated AMP instance where the maximal budget spent by day is exactly B. Reciprocally, in any solution of the associated AMP instance, the total value spent by day is B and each meal corresponds to a distinct s_i. This implies that the S_j as defined in (a) will constitute a cover and the sum of each S_j will be equal to B. For concluding the proof, we must now ensure that the reduction is done in polynomial time.

Given an instance of 3-partition, the construction of the associated AMP can be done in $O(l)$. Once, the instance is solved, the construction of S_j can be done in $O(l^2)$. Indeed, it suffices to loop over each $\sigma(d_j, t_u, m_i)$. Consequently, we have a polynomial time reduction.