Estimating the Nutrient Content of Commercial Foods from their Label Using Numerical Optimization

Jieun Kim(✉) and Mireille Boutin

Purdue University, West Lafayette, IN 47907, USA
`mboutin@purdue.edu`

Abstract. We propose a method for automatically estimating the amount of a given nutrient contained in a commercial food. The method applies when no part of any ingredient is removed in the preparation process. First, we automatically bound the amount of each ingredient used to prepare the food using the information provided on its label (Ingredient list and Nutrition Facts Label) along with the nutrition information for at least some of the ingredients. Using these bounds (minimum and maximum amount for each ingredient), we obtain an initial set of bounds (minimum and maximum amount) for the nutrient considered. We then utilize the Simplex algorithm to refine these bounds on the nutrient content. Our motivating application is the management of medical diets that require keeping track of certain nutrients such as phenylalanine (Phe) in the case of the inherited metabolic disease phenylketonuria (PKU). To test our method, we used it to estimate the Phe content of 25 commercial foods. In a majority of cases (17/25), the bounds obtained were within 10.4mg of each other and thus our method provided a very accurate estimate (±5.2mg) for the Phe content of the foods.

1 Introduction

Some medical diets require keeping track of one's intake of certain nutrients. In order to do this, patients need to have access to the nutritional information for the food they consume. While many nutrients are listed on the Nutrition Facts Label of commercial foods, the information provided is not complete. Indeed, not all nutrients are listed on the label, and the content for the ones that are listed is rounded. Being able to automatically determine the amount of a nutrient contained in the food would thus be helpful for these patients.

Our motivating application is the management of inborn errors of metabolism, the most common of which is phenylketonuria (PKU). PKU is characterized by an inability to metabolize Phenylalanine (Phe) [2], which leads to an abnormal accumulation of Phe in the patient's blood. High blood Phe levels affect the patient's neurological system; they are especially detrimental to the intellectual growth of infants. Thus, one of the main goals in the clinical management of PKU is to maintain low blood Phe levels. For most patients, this is achieved by following a strict Phe-restricted diet [5] which requires measuring

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 309–316, 2015.
DOI: 10.1007/978-3-319-23222-5_38

the food consumed and multiplying the amount by a food specific Phe ratio. Unfortunately, the Phe content of commercial foods is not listed on the Nutrition Facts Label, and so patients must obtain the Phe ratios from a food list (e.g., [4,6,7]). As these databases only list a limited number of foods, alternative methods for finding the Phe content of foods would be desirable.

In this paper, we propose to estimate the content of a given nutrient such as Phe by obtaining a minimum bound and a maximum bound for the nutrient amount contained in the food. To do this, we use the food label (Nutrition Fact Label and Ingredient list), along with the USDA Food Database [7] (USDA database).

From the food label, we get the serving size x and the n ingredients used in the recipe. Let A_i denote the weight (in grams) of ingredient i, for $i = 1, \ldots, n$. Since the ingredients are listed in decreasing order of weight, we have $A_i \geq A_{i+1}$. If no part of any ingredient is removed in the preparation process, we thus have

$$x \geq A_1 \geq A_2 \geq \ldots \geq A_n > 0, \tag{1}$$

$$A_1 + A_2 + \ldots + A_n = x. \tag{2}$$

The food label gives us the rounded content y^{nut} of many nutrients. Let Δ^{nut} be the rounding error for the content of nutrient "nut". We can look for the amount y_i^{nut} of nutrient "nut" in one gram of ingredient i in the USDA database. If no part of any ingredient is removed in the preparation process, we have

$$y^{nut} - \Delta^{nut} \leq \sum_{i=1}^{n} y_i^{nut} A_i \leq y^{nut} + \Delta^{nut}. \tag{3}$$

In a preliminary version of this work [3], we proposed an iterative method for finding bounds, $A_{i_{min}}$ and $A_{i_{max}}$, for each ingredient amount A_i, which is applicable even if the nutrient data for some of the ingredients is missing. The bounds obtained this way yield a first set of bounds for the amount of the considered nutrient (e.g., Phe) contained in the food. This is Step 1 of our proposed method for nutrient content estimation, which we describe in Section 2. This step requires prior knowledge of the amount of the considered nutrient (e.g., Phe) for each ingredient. For example, when trying to estimate the Phe content of the food, then the Phe contents for all the ingredients must be known. Since many ingredients not listed in the USDA database clearly do not contain a significant amount of proteins (e.g., food coloring, natural flavor, etc.) and thus can be considered free of Phe, this is a reasonable assumption.

In Step 2 of our method, we make use of the Simplex algorithm in order to further narrow the interval of bounds for the nutrient content. This step is described in Section 3. Our method (Step 1 and Step 2) is applied to the problem of approximating the ingredient amounts and estimating the Phe content of various commercial foods in Section 4. We conclude in Section 5.

2 Step 1: Nutrient Content Estimation Using Approximate Ingredient Amounts

If we knew A_i, the amount of ingredient i, along with p_i, the number of milligrams of a given nutrient per gram of ingredient i, then $p_i A_i$ would be the nutrient contributed by ingredient i, and the total given nutrient in the food would be $\Sigma_{i=1}^{n} p_i A_i$. Therefore, we have the following bounds for the nutrient content (NUT),

$$\Sigma_{i=1}^{n} p_i A_{i_{min}} \leq NUT \leq \Sigma_{i=1}^{n} p_i A_{i_{max}}. \tag{4}$$

An approximate inverse recipe method to estimate a minimal and maximal bound for each ingredient ($A_{i_{min}}$, $A_{i_{max}}$), previously proposed in [3]. We briefly describe the method in Section 2.1. These estimates shall then be put into Equation (4) to obtain a first set of bounds for the content of the considered nutrient NUT.

2.1 Estimating Ingredient Amounts Based on an Approximate Inverse Recipe Method

To obtain an initial range estimate for each ingredient i, Procedure 1 is applied. Then, the initial bounds $A_{i_{min}} \leq A_i \leq A_{i_{max}}$ are narrowed further using Procedure 2.

Further refinement can be obtained using Procedures 3 and Procedure 4. We apply Procedure 3 to increase the minimal bound and Procedure 4 to decrease the maximal bound. Note that the minimal bound can only be refined if y_i^{nut} is known for all i. Otherwise, the bound remains as it is. This is not the case for the maximal bound.

Procedure 1. Initial bound

```
$A_{1_{min}} \leftarrow \frac{x}{n}$,    $A_{1_{max}} \leftarrow x$
for $i = 2$ to $n$ do
    $A_{i_{min}} \leftarrow 0$,    $A_{i_{max}} \leftarrow \frac{x}{i}$
for given nutrient with content $y^{nut}$ do
    if $y_1^{nut} \neq 0$ then
        $A_{1_{max}} \leftarrow min(A_{1_{max}}, \frac{y^{nut}+\Delta^{nut}}{y_1^{nut}-\Delta_1^{nut}})$
    for $i = 2$ to $n$ do
        if $y_i \neq 0$ then
            $A_{i_{max}} \leftarrow min(A_{i_{max}}, A_{i-1_{max}}, \frac{y^{nut}+\Delta^{nut}}{y_i^{nut}-\Delta_i^{nut}})$
        else
            $A_{i_{max}} \leftarrow min(A_{i_{max}}, A_{i-1_{max}})$
```

To estimate the A_i's, we first select a set of nutrients that are listed on the Nutrition Facts Label (e.g., carbohydrates, sodium, protein, etc.). We then apply Procedure 1 (running over all selected nutrients), followed by Procedure 2.

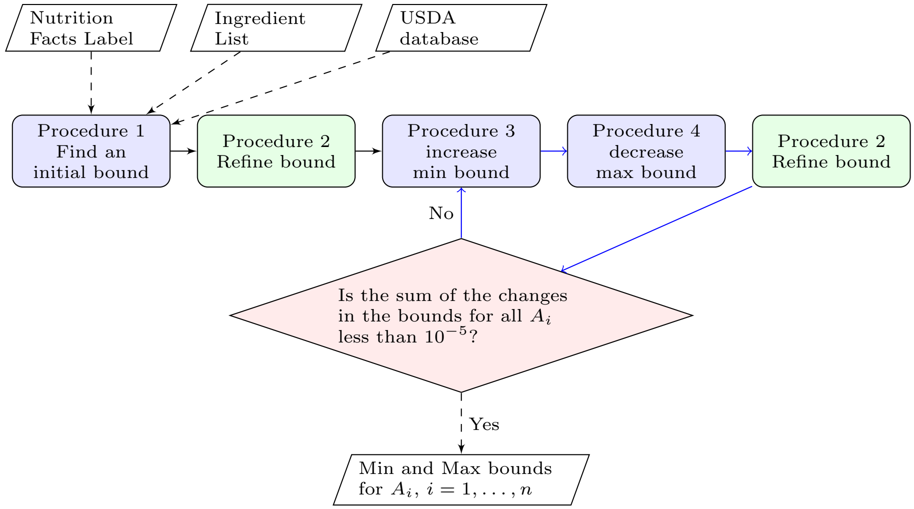

Fig. 1. Schematic Diagram of Proposed Method to Estimate the Ingredient Amounts

Procedure 2. Refining bound

```
for $i = 1$ to $n$ do
    $A_{i_{min}} \leftarrow max(A_{i_{min}}, x - \Sigma_{j=1,j\neq i}^{n} A_{j_{max}})$
    $A_{i_{max}} \leftarrow min(A_{i_{max}}, x - \Sigma_{j=1,j\neq i}^{n} A_{j_{min}})$
```

Procedure 3. To increase the minimal bound

```
for given nutrient with content $y^{nut}$ such that $y_k^{nut}$ exists $\forall k$ do
    if $y_n^{nut} \neq 0$ then
        $A_{n_{min}} \leftarrow max(A_{n_{min}}, \frac{(y^{nut}-\Delta^{nut})-\Sigma_{k=1}^{n-1}(y_k^{nut}+\Delta_k^{nut})A_{k_{max}}}{y_n^{nut}+\Delta_n^{nut}})$
    for $i = n-1$ to 1 do
        if $y_i^{nut} \neq 0$ then
            $A_{i_{min}} \leftarrow max(A_{i_{min}}, A_{i+1_{min}}, \frac{(y^{nut}-\Delta^{nut})-\Sigma_{k=1,k\neq i}^{n}(y_k^{nut}+\Delta_k^{nut})A_{k_{max}}}{y_i^{nut}+\Delta_i^{nut}})$
```

Procedure 4. To decrease the maximal bound

```
for given nutrient with content $y^{nut}$ do
    for $k = 1$ to $n$ do
        if $y_k^{nut}$ does not exist then
            $y_k^{nut} \leftarrow 0, \quad \Delta_k^{nut} \leftarrow 0$
    if $y_1^{nut} \neq 0$ then
        $A_{1_{max}} \leftarrow min(A_{1_{max}}, \frac{(y^{nut}+\Delta^{nut})-\Sigma_{k=2}^{n}(y_k^{nut}-\Delta_k^{nut})A_{k_{min}}}{y_1^{nut}-\Delta_1^{nut}})$
    for $i = 2$ to $n$ do
        if $y_i^{nut} \neq 0$ then
            $A_{i_{max}} \leftarrow min(A_{i_{max}}, A_{i-1_{max}}, \frac{(y^{nut}+\Delta^{nut})-\Sigma_{k=1,k\neq i}^{n}(y_k^{nut}-\Delta_k^{nut})A_{k_{min}}}{y_i^{nut}-\Delta_i^{nut}})$
```

After that, we keep repeating Procedure 3 and Procedure 4 (running over all selected nutrients), followed by Procedure 2, until our estimates change by less than 10^{-5} between consecutive repetitions. This is illustrated in Figure 1.

The accuracy of our method depends on the food considered and can vary from one ingredient to the next. However it is not necessary that all ingredient amounts be precisely estimated in order to get a good estimate on the content of the query nutrient NUT.

3 Step 2: Nutrient Content Estimate Refinement Using Simplex Algorithm

Observe that the bounds obtained using Equation (4) correspond to ingredient amounts that can violate Equation (2). More specifically, neither $\sum_{i=1}^{n} A_{i_{min}}$ nor $\sum_{i=1}^{n} A_{i_{max}}$ equal to a serving size x in general. This indicates that it should be possible to further refine the content estimate obtained in Step 1. We propose to do this using the Simplex algorithm [1] which is a well-known linear programming tool. The Simplex algorithm first finds an initial feasible solution in Phase I. Then, in Phase II, it moves along the edges of the polytope defined by the constraints while evaluating the cost until it reaches an extreme value. In the case of the nutrient content estimation problem, the cost function is the summation of the nutrient content coming from each ingredient, the nutrient content (NUT), $\sum_{i=1}^{n} p_i A_i$.

The linear constraints of the nutrient content estimation problem are defined by Definition 1. A feasible solution to this problem, a set of A_i, must be selected from the obtained minimal and maximal bounds for each ingredient and should satisfy our assumption that there is no loss of any ingredient. Therefore, the linear constraints are composed of Equation (1), Equation (2), and the obtained ranges of ingredient amounts in Section 2.1.

Definition 1. Nutrient content estimate using the Simplex algorithm

---

minimize, maximize $\sum_{i=1}^{n} p_i A_i$ *where*

$$\begin{cases} \sum_{i=1}^{n} A_i = x \\ x \geq A_1 \geq A_2 \geq \cdots \geq A_n > 0 \\ A_{i_{min}} \leq A_i \leq A_{i_{max}}, \quad i = 1, \ldots, n \end{cases}$$

Since all constraints are equalities, any feasible solutions satisfying the constraints are points on the edges of a (n-1)-dimensional polytope. Hence, once an initial feasible point is found from Phase I of the Simplex algorithm, in Phase II, we look through the extreme points of the polytope until the cost at any adjacent points of an extreme point does not decrease anymore. The cost at the point becomes the minimum of the nutrient content for a serving size x gram of a food. Similarly, once the cost function does not increase anymore, we set the maximum bound for the nutrient content to the value of the cost function.

4 Numerical Experiment: Application to Phenylalanine (Phe) Content Estimation

We experimented with our method to estimate the Phenylalanine (Phe) content for 25 commercial food items. The results are shown in Table 1. The table presents the minimum and maximum bounds for each food obtained by our method using six nutrients (protein, sodium, calories, carbohydrates, fat, and cholesterol). Both the results after Step 1 (column 4) and Step 2 (column 5) are given in order to see the improvement resulting from performing Step 2.

For comparison, the Phe data from two databases, USDA database [7] and a low-protein food database [6], are written in the first and second column of Table 1. When there exists no data related to the item from that database, we indicated the case with ‘N/A’. As we expected, only a part of the food items considered has Phe data in the USDA database (6/25) or Phe data in the low-protein food database (14/25). Furthermore, some of the data listed in our table may be inexact as we were unable to find the specific brand of product considered and used a generic version instead. For example, the Phe content for Tomato soup specifically from Campbell company is not presented in the USDA database while the USDA database contains the Phe content of Tomato soup for any brand.

In contrast, Step 1 of our approximate method was able to provide bounds for the Phe content of all targeted food items, as shown in the fourth column of Table 1 where the estimated minimum and maximum values for the Phe content are written in parenthesis. The range between the minimum and the maximum bounds was less than 10mg for 16 food items, and less than 25mg for 19 food items. The estimated bounds for the Phe content were within no more than 3 mg from at least one of the databases for 22 items, which is 88% of the 25 foods considered. In the case of butter, rice krispies cereal and waffles, our range exclude the Phe value from both databases. This is most likely due to the violation of our assumption that no part of any ingredient is discarded during the preparation process. For example, there is considerable drying in the preparation of cereal, and liquid (whey) is discarded in the preparation of butter.

After Step 2, the interval between bounds for the Phe content narrowed significantly more in 10 cases (see the fifth column of Table 1). Step 2 narrowed the range of the estimated Phe bounds for one serving of Salsa Sauce from 24.68mg to 10.33mg. In the case of garlic mashed potatoes and sweet potato tot, the ranges of the estimated bounds for Phe content decreased to the values less than half of the ranges after Step 1. Moreover, the largest range between the minimum and maximum bounds after Step 2 became 54.76mg, less than one third of the highest range after Step 1 (165.61mg). The Simplex algorithm in Step 2 could not find an initial feasible solution for 9 items; these are denoted by ‘DNE^{c}’ in the table. This could be because an ingredient used to prepare the food did not coincide with the ingredient listed in the USDA database. Another inconsistency could have occurred because we neglected ingredients with negligible amounts for which the USDA database did not provide any data. However, even though we could not improve the bounds for the Phe content any further for these 9

Table 1. Comparison of phenylalanine content estimates obtained with our methods and two food composition databases.

Description(serving size)	USDA database[7]	low-protein food database[6]	After Step 1	After Step 2
Carr's Whole Wheat Crackers(17g)	81.6mg	75mg	(53.61mg, 85.11mg)	(53.61mg, 85.11mg)
Ketchup(17g)	4.42mg	10.2mg	(0.70mg, 7.09mg)	(1.20mg, 6.57mg)
KIT KAT Milk Chocolate(42g)	113.4mg	131.86mg	(129.56mg, 238.91mg)	(144.27mg, 191.53mg)
Campbell's Tomato soup(122g)	68.32mg[a]	66.90mg	(33.21mg, 102.91mg)	(40.69mg, 95.45mg)
Cheerios Cereal(28g)	175.84mg	165mg	(179.86mg, 180.51mg)	DNE[c]
Rice Krispies Cereal(33g)	116.82mg	107mg	(91.54mg, 94.80mg)	DNE[c]
Enchilada Sauce(60g)	N/A	6mg	(0.41mg, 35.69mg)	(0.41mg, 34.14mg)
Eggo waffle(70g)	N/A	238mg	(196.26mg, 216.35mg)	(196.26mg, 216.35mg)
Garlic chili pepper sauce(9g)	N/A	1.93mg	(1.37mg, 6.96mg)	(2.65mg, 5.27mg)
Salsa sauce(30g)	N/A	11mg	(1.53mg, 26.21mg)	(7.90mg, 18.23mg)
Garlic mashed potatoes(124g)	N/A	N/A[b]	(56.89mg, 222.50mg)	(139.51mg, 162.23mg)
Butter with Canola Oil(14g)	N/A	6mg	(11.88mg, 17.66mg)	(12.06mg, 17.66mg)
Go-Gurt(64g)	N/A	120mg	(116.38mg, 120.95mg)	DNE[c]
Jell-O Gelatin Snacks(98g)	N/A	23.76mg	(10.01mg, 30.44mg)	(10.01mg, 30.44mg)
Marshmallow Peeps(42g)	N/A	21mg	(19.17mg, 23.56mg)	DNE[c]
Ore-Ida French fries(84g)	N/A	76mg	(77.64mg, 78.77mg)	(77.64mg, 78.76mg)
Spicy Brown Mustard(5g)	N/A	8mg	(9.87mg, 10.35mg)	(10.11mg, 10.16mg)
Starburst Fruit Chews(40g)	N/A	5.42mg	(0.00mg, 4.48mg)	DNE[c]
Vinaigrette Balsamic Dressing(31g)	N/A	3mg	(0.00mg, 5.53mg)	(0.00mg, 5.53mg)
Yoplait Original Strawberry(170g)	N/A	284.67mg	(287.11mg, 291.08mg)	DNE[c]
ALTOIDS peppermint(2g)	N/A	N/A	(0.43mg, 4.22mg)	DNE[c]
Jell-O Cheesecake Pudding(26g)	N/A	N/A	(0.91mg, 0.98mg)	DNE[c]
Sweet potato Tot(85g)	N/A	N/A	(54.87mg, 113.77mg)	(71.91mg, 95.82mg)
Taco Shells(32g)	N/A	N/A	(36.69mg, 38.31mg)	(36.69mg, 38.31mg)
Vanilla bean Ice cream(87g)	N/A	N/A	(206.87mg, 211.09mg)	DNE[c]

[a] Any brand Tomato soup, condensed. Not Campbell's product.
[b] Database has a value, but with a different protein content.
[c] Simplex algorithm could not find a solution

items after Step 2, notice that the bounds after Step 1 in these cases were already very close to each other, with a difference of less than 5mg per serving size.

5 Summary and Conclusions

The Food Safety and Inspection Service of the USDA (United States Department of Agriculture) mandates food companies to label their products with an Ingredient list and a Nutrition Facts Label. This information is important, but incomplete. Indeed, some nutrients such as Phenylalanine (Phe) are not listed on the label. This is problematic for individuals with inherited metabolic disorders such as PKU who must carefully monitor their Phe intake. In an attempt to help these individuals to manage their medical diets, we proposed a method for estimating the content of a given nutrient automatically from the food label information. We assume that no part of any ingredient is removed while preparing a food. This gives two constraints: the sum of each ingredient content equals to a serving size for the food and the weighted sum of a nutrient content for one gram of each ingredient equals to the nutrient content for one serving of the food. We also use the fact that the ingredients are listed in decreasing amounts (per weight). The proposed method is applicable even if the nutrient content of some of the ingredients is not fully known.

We applied our method to the problem of Phe content estimation. Our approach finds bounds for the Phe content of a food based on the estimated ingredient amounts in Step 1. Step 2 refines the results using linear programming (Simplex algorithm). We showed our results for various commercial foods in Table 1. The intervals between the estimated bounds for the Phe content after Step 2 were within 10.4mg for 17 items and within 24mg for 21 items out of the 25 foods considered. In contrast, the intervals were within 10mg for 16 items and within 25mg for 19 items after Step 1.

While two current databases did not contain Phe data for all the food we considered, our method provided a Phe content estimate for all of them. Hence, we believe that our work provides a useful tool to help PKU patients manage their diet. Moreover, our method can be used to estimate other nutrient contents, or to increase the precision of the nutrient content listed on the Nutrition Facts Label. So it should be helpful in managing other diets as well.

References

1. Benhamadou, M.: On the simplex algorithm algorithm revised form. Advances in Engineering Software **33**(11), 769–777 (2002)
2. Huttenlocher, P.R.: The neuropathology of phenylketonuria: human and animal studies. European Journal of Pediatrics **159**(2), S102–S106 (2000)
3. Kim, J., Boutin, M.: An approximate inverse recipe method with application to automatic food analysis. In: 2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), pp. 32–39. IEEE (2014)
4. Kim, J., Boutin, M.: A list of phenylalanine to protein ratios for common foods. ECE Technical Reports. Paper 456 (2014). http://docs.lib.purdue.edu/ecetr/456
5. National Institute of Health: Consensus development conference statement. Phenylketonuria: Screening and Management (October 16–18, 2000). http://www.nichd.nih.gov/publications/pubs/pku/sub3.cfm
6. Schuett, V.E.: Low Protein Food List for PKU, 3rd edn (2010)
7. U.S. Department of Agriculture, Agricultural Research Service: USDA national nutrient database for standard reference, release 25. Nutrient Data Laboratory Home Page (2012). http://www.ars.usda.gov/ba/bhnrc/ndl