A diet expert system utilizing linear programming models in a rule-based inference engine

Annette van der Merwe 1,* , Hennie Kruger 1 and Tjaart Steyn 1
1North-West University, Potchefstroom Campus, Potchefstroom, South Africa

Abstract—Linear programming is commonly used for solving complex problems in various fields, including dietetics. Expert systems use expertise and inference procedures to solve problems that require advanced expert knowledge and are also applied to health related problems. Over the years many variations and facets of the diet problem and other related problems have been solved by means of linear programming techniques as well as expert systems. In this research, an expert system was created for the purpose of solving multiple facets of the diet problem, by creating a rule-based inference engine consisting of goal programming- and multi-objective linear programming models. The program was successfully applied to cases specific to South African teenage girls, which were obtained through system development. The resulting system compiles an eating-plan for a girl that conforms to the nutritional requirements of a healthy diet, includes the personal food preferences of the girl, and consists of food items that result in the lowest total cost. The system also allows prioritization of the food preference and least cost factors by means of weighted priorities.

Introduction

Studies have shown that obesity and other eating disorders are on the increase in South Africa, especially among female adolescents as discussed in Kimani-Murage (2013), Reddy et al (2012), and Monyeki et al (2012). One of the reasons for this, as addressed by Schönfeldt et al (2010), is inefficient education regarding eating habits appropriate for individuals and the inability to gain the necessary education due to financial constraints. People also tend to deviate from eating plans to cater for personal food preferences. Neumark-Sziffer (2009) presented five main proposals for preventing obesity and related eating disorders among girls, which include eating healthily rather than following diets, adopting a positive body image, having meals with their families instead of their friends, taking part in physical activities, and involving the families of overweight teenagers when addressing weight related problems. According to Temin et al (2010), the problem could also be addressed by effecting low cost changes in the health system which would make the whole sector more responsive to their needs. Finding an ideal solution to the problem means having to consider not only the environment in which these girls exist, but also reflecting on previous work done to solve similar or related problems. Linear programming has been utilized in many different applications for the purpose of solving diet related problems.

Artificial intelligence or more specifically, expert systems, have also been developed to solve either meal planning problems or health related problems. Many free software solutions exist that can assist people in selecting healthy foods to include in their diets. However, when developing a healthy diet for an adolescent, Whitney and Rolfes (2008) considers factors like gender, age, height, weight, activity level and some values obtained through calculations, using exactly these factors. The purpose of this research was to establish a software solution to the diet problem specifically related to South African adolescent girls by developing an expert system consisting of a linear programming model that represents the rule-based inference engine.

The next section will briefly discuss the evolution of the diet problem and how linear programming has been used to solve general and special cases in dietetics, followed by an outline of some expert systems applications which were also used for health related problems. Then the layout and planning of the system is shown and the components briefly explained. This is followed by the formulation of the goal- and multi-objective linear programming models which constitute the inference engine and knowledge base, after which the evaluation and results obtained through experiments are briefly discussed. The paper finishes with some concluding remarks on the incorporation of linear programming models in the inference engine and how the expert systems- and linear programming fields benefit from the research conducted.

* Correspondence: Annette van der Merwe, North-West University, Potchefstroom Campus, 11 Hoffman Street, Potchefstroom 2531 South Africa
E-mail: annette.vandermerwe@nwu.ac.za
Related Work

Linear programming is commonly used for solving the diet problem which generally involves the optimization of an objective function describing the total cost or the nutrient contribution of a diet. It originated in 1947 when George Dantzig started to develop a method that would help him formulate a diet consisting of all the nutrients necessary to sustain a healthy body (Dantzig, 1990). His work also cemented the concept of ‘the diet problem’ as a means to explain linear programming concepts such as the well-known simplex method, also studied by Rolf (2013). Linear programming and related extensions are now used widely for solving problems that contain many variables and restrictions.

Ever since the start of linear programming, researchers have utilized it in various ways to solve different adaptations of diet or health related problems. Anderson and Earle (1983) developed a linear programming and goal programming approach that minimized the shortfall and excess of the macronutrients provided by foods that the World Health Organization distributed to Third World Countries. Their work helped to balance the nutrient levels required by people in starving countries. In 1989 the Food and Nutrition Board drafted a set of nutritional guidelines which are known as the Recommended Daily Allowances. This list has been used as general dieting guidelines ever since.

Another application of linear programming was performed by Sklan and Dariel (1993) when they used mixed-integer linear programming to eliminate the repetition of food types in a diet by constraining the number of times certain food items appeared in a meal plan. They used the simplex method by applying additional integer and bound restrictions on the diet and their method could solve 800 equations with 4000 variables. This resulted in a repetitive, iterative method which provided several solutions to the same problem, from which the optimal solution was determined much like the branch-and-bound method as also discussed by Vanderbriei (2008).

More recently, Ferguson et al (2004) implemented a four stage linear programming model to set food based dietary guidelines for people who live in Malawi. The solution consisted of a goal programming model which optimized the energy contribution and also included food items most popular in the country. The work successfully resulted in a nutritionally balanced diet which incorporated food items frequently consumed in a specific region.

Linear programming applications do not only apply to human diets, but have also been used to solve animal rationing problems as used by Aldesait et al (2012). They constructed a linear programming model that successfully created a least cost fodder ration combination with the aim of optimizing the fattening of calves at different age stages. With the ever increasing rate at which technology develops, it is only natural that linear programming models be applied in computerized systems. Examples of such systems that have been created for diet or health related issues include the work of Bassham et al (1984) which assisted dieticians working in hospitals to analyze the diets of patients. Soden and Fletcher (1992) developed a program that would modify an existing diet to conform to the nutritional requirements of a certain individual. Cadenas et al (2004) introduced a computer program, SACRA, which could be applied to develop accurate feed-mixes for livestock in Argentina, while taking into account that all animals did not eat similar daily amounts.

Kahraman and Seven (2005) presented a computer system that utilized the branch-and-bound method to minimize a diet in terms of cost, while attempting to include most of a certain individual’s food preferences. Frega et al (2012) developed a program that could be used to evaluate the average dietary needs in a typical Mozambican household and present a healthy diet for such a family. Although the system provided feasible solutions regarding dietary constraints and requirements, the resulting diets were not generally very affordable.

Another interesting implementation of linear programming is seen in the work of Mamat et al (2011). Their model incorporated fuzzy numbers to address fluctuating food prices by considering maximum and minimum price levels for food items. The resulting program could successfully develop low cost, -carbohydrate, and -fat diets. While linear programming and related extensions are widely used to solve diet and health related problems, the use of these models are increasingly being combined with alternative technologies like expert systems. Giarratano and Riley (2005) state that an expert system utilizes expert knowledge and inference procedures to solve problems that require advanced human expertise. Many expert systems exist that aim to solve specific elements of diet- and health related problems, including:

- Sterling et al (1996) who conducted a study to determine the importance of common sense knowledge in the construction of a meal plan by creating four different expert systems which were used for menu planning. Their work resulted in a program that used layered networks to establish relationships between parts of a created menu.
- Lo et al (2011) used a medical tourism expert system that recommended natural medicine alternatives to people by creating health profiles for the users.
- Ting et al (2010) created the CASESIAN expert system which helped to facilitate the sharing of clinical insights and practice experience in the healthcare industry.
Berry et al (1995) developed an expert system called OPADE, which aimed at improving the drug prescribing behavior of doctors.

In this research, an expert system was created with a rule-based inference engine consisting of linear programming models which generate an eating plan conforming to the nutritional requirements and personal food preferences of a South African teenage girl.

**Expert system layout**

An expert system is generally created according to a knowledge engineering process. This process consists of the planning, knowledge definition, knowledge design, and code and checkout stages. As part of the planning stage, a preliminary functional layout was created that consisted of the main components needed for the system. These components include the user interface, working memory, inference engine and the knowledge base.

The components were constructed in the form of a Microsoft Excel Workbook:

- The user interface consists of an interactive combination of spreadsheets and forms which control the process of obtaining the necessary information from the user and providing output in the form of an eating plan.
- The working memory is the dynamic memory of the system which temporarily holds the information entered by the user and is responsible for compiling the data into the format required by the inference engine. For this purpose the system utilizes Visual Basic for Applications (VBA) code, as discussed in detail in Albright (2012).
- The inference engine can generally be described as the brain of the system and usually contains a set of production rules. A production rule can be described as an IF-THEN operation, where an action is typically performed (THEN-portion of the rule) only after a condition is satisfied (IF-portion of the rule). The function of the inference engine is to use the data acquired from the user to make certain conclusions, in this case, regarding the food items to include into a diet plan. The inference engine in this system, implements two interactive linear programming models by means of MS Excel Solver that applies the data acquired from the user, to the knowledge base in order to develop an eating plan. The inference engine will be discussed in more detail in the next section.
- The knowledge base consists of the knowledge acquired from an expert in dietetics which in this application represents the constraints relevant to an eating plan for a certain individual. The knowledge base consists of pre-compiled lists of foods native to South Africa, grouped according to the proportions of carbohydrate (CHO), protein and fat they provide. These lists are called exchange groups which are dietary tools used to assist a dietician to organize foods according to their macronutrient content. The exchange groups include fruit, vegetable, starch, dairy, fat, and meat and legumes. A dietician typically selects a certain number of foods, or exchanges, from each exchange group for incorporation into the final eating plan.

The basic components of the expert system are shown in Figure 1.

---

**Fig. 1. Layout of the expert system components**
During knowledge definition, several sources need to be explored in order to establish a comprehensive knowledge base required for the system. Sources consulted in this research include books, academic study material, human experts (lecturers in dietetics and practicing dieticians), electronic and internet resources, and journal articles and publications. A practicing dietician was asked to consult with six volunteer female teenagers in the Northwest province of South Africa and develop an eating plan for each individual, Rankin (2012). The volunteers were chosen to represent most cases based on varying activity levels and eating perceptions so as to minimize duplicated sessions. The purpose of these sessions was to experience how a dietician applies theoretical processes in practice so that the rules can be created for the system.

The dietician was asked to create a healthy diet for each of the volunteers and the sessions were recorded and transcribed. During these sessions it became evident that the dietician intuitively aimed to select cheap food options for a diet so as to keep the total cost a minimum, but used no formal calculations for this purpose. The dietician also emphasized that some difficulty is usually experienced in varying the food choices which are usually selected habitually. The girls were asked general questions related to food preferences and eating habits and a balanced eating plan was provided to each of them. These eating plans were also used to evaluate the created expert system. Additionally, lists of food types typically consumed in high school dormitories in Potchefstroom were compiled as part of a related research study by Linde et al (2012). These foods were incorporated into the system for determining the preference levels of a user for each food type.

Fig. 2. Function of the expert system
The knowledge definition stage resulted in the following functional requirements for the system:

- Calculating the daily energy requirement.
- Presenting food lists to the user for food preference acquisition.
- Optimizing user food preference and total cost by solving the linear programming models in the inference engine.
- Providing output to the user.

The function of the expert system is shown in Figure 2. The goal- and multi-objective linear programming models interact to perform the functions of the inference engine by solving the models and the knowledge base by incorporating the rules as constraints.

General information required from the user includes factors like age, gender, weight, height, and activity level. The program starts by calculating the basal metabolic rate (BMR) using the Harris-Benedict formulas, which according to Nienaber-Rousseau et al (2012) are the following:

For males: \[ BMR = (66.5 + 13.8m + 5h - 6.8a) \times 4.2 \]

For females: \[ BMR = (655 + 9.6m + 1.8h - 4.7a) \times 4.2 \]

where \( m \) denotes the mass of the person in kg, \( h \) denotes the height of the person in cm, and \( a \) denotes the age of the person in years.

The BMR, activity level and energy required for growth based on the person’s age are then used to determine the daily required energy (RE) for the individual. The user is required to enter the preferred distribution percentages of the RE into the three macronutrients protein, carbohydrates (CHO) and fat. The specific amounts of each macronutrient are then calculated from the RE, using the following formulas, where 17kJ of protein, 17kJ of CHO, and 38kJ of fat weighs 1g.

protein \( (g) = \frac{\text{\% protein} \times \text{total daily energy requirement (kJ)}}{17 (kJ/g)} \)

CHO \( (g) = \frac{\text{\% CHO} \times \text{total daily energy requirement (kJ)}}{17 (kJ/g)} \)

fat \( (g) = \frac{\text{\% fat} \times \text{total daily energy requirement (kJ)}}{38 (kJ/g)} \)

### Goal programming and multi-objective linear programming inference engine and knowledge base

The inference engine and part of the knowledge base of the expert system was implemented using a combination of a goal programming model and a multi-objective linear programming model. The goal programming model was implemented to determine the number of exchanges to be chosen from each exchange group so that the macronutrient distribution is as close as possible to the percentages either entered by the user or chosen by the dietician. The model is formulated as follows:

**Minimise** \[ \sum_{i=1}^{k} (u_i + v_i) \]

**Subject to**

\[ \sum_{j=1}^{n} a_{ij}x_j + u_i - v_i = M_i \quad (i = 1, ..., k) \]

\[ l_j \leq x_j \leq b_j \quad (j = 1, ..., n) \]

\[ u_i, v_i \geq 0 \quad (i = 1, ..., k) \]

\[ x_j, l_j, b_j \text{ integer} \quad \forall j \]

where \( u_i \) and \( v_i \) represent the positive and negative deviations of the \( i^{th} \) macronutrient; \( l_j \) and \( b_j \) represent the lower- and upper bounds for exchange group \( j \); \( M_i \) represents the weight amount in grams required from the \( i^{th} \) macronutrient; \( k \) is the number of macronutrients; and \( a_{ij} \) represents the contribution of exchange group \( j \) to the \( i^{th} \) macronutrient.

The upper and lower bounds for each exchange group are constrained to integer values and are determined using the food guide pyramid as developed by the United States Department of Agriculture (http://fnic.nal.usda.gov/dietary-guidance/myplatefood-pyramid-resources, 15 December 2013). This is an internationally accepted set of dietary guidelines. This model provides a concrete integer number that indicates how many food selections should be made from each of the exchange lists. The next phase of the program uses a multi-objective linear programming model to select foods from each exchange group. The model is formulated as follows:
Minimise $Q$
Subject to
\[
\frac{w_{\text{pref}}(\sum_{j=1}^{n} p_{j} x_{j} - t_{\text{pref}})}{t_{\text{pref}}} \leq Q
\]
\[
\frac{w_{\text{cost}}(\sum_{j=1}^{n} s_{j} x_{j} - t_{\text{cost}})}{t_{\text{cost}}} \leq Q
\]
\[
\sum_{j=1}^{n} x_{j} = N
\]
\[
x_{j} = 0 \text{ or } 1
\]

where $N$ denotes the number of selections to be made; $x_{j} = 1$ if food $j$ is selected and 0 otherwise; $p_{j}$ depicts the user’s preference for food item $x_{j}$; $s_{j}$ represents the cost of food item $x_{j}$; $w_{\text{pref}}$ represents the weight as selected by the user for the food preference as priority; $w_{\text{cost}}$ represents the weight for cost; $t_{\text{pref}}$ represents the optimum target for preference rating; and $t_{\text{cost}}$ represents the optimum target for cost.

Because many similar products exist, additional constraints were added to the model. This would ensure that for example when a muffin is selected for a diet, a crumpet is not also chosen. These constraints are incorporated in the model in the following form:

\[
\sum_{k=1}^{p} y_{k} = 0 \text{ or } 1
\]

where $p$ denotes the number of similar or like items of which only one must be selected, and $y_{k}$ is the $k^{th}$ item in the list of similar items.

This model determines, for each of the exchange groups, a selection of food items that provides the optimal level of cost and user food preference. It uses the number of foods to select from each of the exchange groups, calculated in the goal programming model, and the food preferences of the user as inputs. In effect, the model represents the rules that would result in the combination of foods with the best preference rating, based on the user’s food preferences.

**Evaluation and results**

The created expert system was evaluated in terms of two criteria: the effect of changing the weighted priorities for cost and preference was inspected; and the system was applied to the six available real-world case studies and the results were compared. For demonstrating the effect that changes in the cost- and preference priority have on the resulting eating plan, an experiment was conducted by applying one of the case studies to the system with varying priority levels. The particular eating plan was created for a 6164 kJ daily energy requirement and the priorities were varied as shown in Table 1. Figure 3 presents the results of the experiment.

**Table 1. Cost versus preference weights used for applying the case study to the system**

<table>
<thead>
<tr>
<th>Application</th>
<th>1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>4&lt;sup&gt;th&lt;/sup&gt;</th>
<th>5&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost weight</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Preference weight</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The results show that when the priority of cost was high, not all the user’s preferred foods appeared in the final eating plan because the system aimed to find a least cost diet. Similarly when the priority for the preference was at the maximum level, the total cost of the resulting diet was disregarded when selecting food items, due to the cost priority of zero. The results are summarized in Table 2, where cost is indicated in R, South African currency.
Fig. 3. Effect of changes in cost versus preference priority on the price of the resulting diet.

Table 2. Statistics for the 6164 kJ eating-plan generated by the expert system

<table>
<thead>
<tr>
<th>Cost</th>
<th>Pref.</th>
<th>Total cost of the diet</th>
<th>Food Preference satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>R 28.76</td>
<td>79.17%</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>R 29.41</td>
<td>82.14%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>R 30.45</td>
<td>82.81%</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>R 32.59</td>
<td>84.72%</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>R 53.14</td>
<td>88.16%</td>
</tr>
</tbody>
</table>

The statistics show that as the priority for cost decreases and that for food preference increases, the price of the resulting diet increases and the food preference satisfaction rises. Further testing was done by using the system to imitate the consultation sessions between the volunteers and the dietician and comparing the resulting eating plans. As six volunteers participated in the sessions, these were the only eating plans available for initial testing. The use of school children for the study meant having to gain parental permission and also bearing the cost of all of the sessions. Comprehensive testing using a larger group is considered. The priorities for the total cost versus user food preference were set to equal weights because they were not considered during the sessions with the dietician. Table 3 shows how the eating plans generated by the system compares to those received from the dietician. For a total of 146 food items used in the system generated plan, the average number of food items similar to those in the dietician’s plans was 68.07%, with 24.97% of food items from the same exchange groups, and 6.96% of foods that were selected from different exchange groups.
Table 3. The number of food matches, group matches and exceptions between the eating-plans

<table>
<thead>
<tr>
<th></th>
<th>Total food items in the plan</th>
<th>% similarity between the plans</th>
<th>% variation provided by the program</th>
<th>% exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volunteer 1</td>
<td>26</td>
<td>76.92%</td>
<td>19.23%</td>
<td>3.85%</td>
</tr>
<tr>
<td>Volunteer 2</td>
<td>26</td>
<td>65.38%</td>
<td>26.92%</td>
<td>7.69%</td>
</tr>
<tr>
<td>Volunteer 3</td>
<td>21</td>
<td>57.14%</td>
<td>33.33%</td>
<td>9.52%</td>
</tr>
<tr>
<td>Volunteer 4</td>
<td>23</td>
<td>60.87%</td>
<td>34.78%</td>
<td>4.35%</td>
</tr>
<tr>
<td>Volunteer 5</td>
<td>24</td>
<td>75.00%</td>
<td>12.50%</td>
<td>12.50%</td>
</tr>
<tr>
<td>Volunteer 6</td>
<td>26</td>
<td>73.08%</td>
<td>23.08%</td>
<td>3.85%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>68.07%</strong></td>
<td><strong>24.97%</strong></td>
<td><strong>6.96%</strong></td>
</tr>
</tbody>
</table>

The volunteers were asked to name their favorite and least favorite foods but most of them only mentioned a few food items meaning that the percentage calculation of the preference satisfaction was done for a very small number of foods when applied to the available cases. The expert system however, presented all of the food items in the respective exchange lists to the user in order to establish her personal preference for each item. Accordingly, the variation in preference percentage levels of some of the sessions when applied to the system was very small.

**Conclusion**

The research provided evidence that expert system technology could be successfully combined with mathematical programming techniques by incorporating linear programming models to perform rule-based inference. The created system acts like an expert in dietetics by performing a basic evaluation, calculating the required daily energy intake for an individual and providing a healthy eating plan. It also has the added advantage of the combination of two powerful solution techniques providing the possibility of solving more than one facet of the diet problem simultaneously. A least cost diet that conforms to the daily energy requirements and consists of food choices preferred by that individual can be generated.

This work offers some valuable contributions to the linear programming and expert systems fields. As far as could be ascertained, using expert system technology in conjunction with integer goal programming as well as multi-objective linear programming to solve the diet problem for a South African individual, has not been performed before. The research presents a successful software solution that automatically generates the constraints required to solve the models through the use of expert system components. The constraints are then used in the linear programming models which represent the rule-based inference engine of the expert system.

Existing systems that utilize linear programming to solve the diet problem generally either minimizes the cost of a diet or provides a diet optimized in terms of food preference. The system developed in this research however, optimizes both of these objectives simultaneously. The system also generates eating plans that are unique to South Africa, as the food lists in the knowledge base have been compiled from South African sources and reflect price information current at the time of coding. According to literature, this is the first time a research study of this nature has been done specifically for South African food types. An alternative method for determining and using the food preference rating of individuals were also implemented successfully by minimizing the inverse of the highest preference levels. This meant that both the cost and preference objectives were minimized. The program also successfully implemented weighted priorities with which the user can determine the level of importance of having a least cost eating plan or rather one that contains as many of the individual’s favorite foods as possible. It further presents the flexibility of allowing the user to choose if the required daily energy intake should be calculated automatically or entered manually. Through the use of multi-objective linear programming models, the system presents a solution to providing some variation in food selections for the resulting eating plan. This could assist a dietician who has trouble varying the food types selected for an eating plan. In terms of future work, the research can be extended by incorporating Pareto-optimal frontiers which could provide the optimal trade-off between the cost and preference level objectives for a diet for a specific individual. Another issue to consider is the static pricing data that was used for calculating the total cost of a diet. Incorporating a dynamic link that automatically updates the prices of food items used in the calculations online, could improve the efficiency of the system considerably an could be considered for future work. Further work could also be considered regarding the micronutrients that food items supply to a diet, instead of only looking at the macronutrients. A study like this however would entail incorporating non-linear programming models into the system.
References

Albright SC (2012). VBA for Modelers: Developing decision support systems with Microsoft (R) Office Excel (R) (4th International ed.). South-Western Cengage Learning, Mason


Nienaber-Rousseau C, Venter CS and Wright HH (2012). Voedingstatus Study guide: VOED 122 PAC. North-West University Potchefstroom Campus, Potchefstroom

Rankin D, (2012). Registered Dietician Personal interview, Potchefstroom


Rolf HL (2013). Finite Mathematics (8th ed.). Cengage Learning, Boston


