Automated Adjustment of Crowdsourced Calorie Estimations for Accurate Food Image Logging

Patrick McAllister¹, Anne Moorhead²*, Raymond Bond¹, Huiru Zheng¹*

School of Computing¹, School of Communication and Media²
Ulster University, Newtownabbey, UK
{mcallister-p2, a.moorhead*, rb.bond, h.zheng*}@ulster.ac.uk

Abstract
Obesity is increasing globally and is a risk factor for many chronic conditions such as heart disease, sleep apnea, type-2 diabetes, and some cancers. Research shows that food logging is beneficial in promoting weight loss. Crowdsourcing has also been used in promoting dietary feedback for food logging. This work investigates the feasibility of crowdsourcing to provide support in accurately determining calories in meal images. Two groups, 1. experts and 2. non-experts, completed a calorie estimation survey consisting of 15 meal images. Descriptive statistics were used to analyse the performance of each group. Collectively, non-experts could determine which meals had larger amounts of calories and analysis showed that meals with greater calories resulted in greater standard deviations of non-expert estimates. Secondary experiments were completed that used crowdsourcing to adjust user calorie estimations using non-expert calorie estimations. Five-fold cross validation was used and results from the calorie adjustment process show a reduced overall mean calorie difference in each fold and the mean error percentage decreased from 40.85% to 25.52% in comparing original mean estimations against adjusted mean estimations. As such, there is credibility in adjusting calorie estimates from a crowd as opposed to simply taking a central measure such as the mean.

1. Introduction

Obesity is a global concern [1] since it increases the risk of acquiring other chronic conditions such as type 2 diabetes, heart disease, and some cancers [2]. Statistics show that in the UK, a large proportion of age groups are overweight and obese. Between 2005 and 2015 study show that there has been a substantial increase in morbidly obese adults of 1.8% to 2.9% and an overall increase from 60.5% to 62.9% in overweight or obese adults for the same period [3]. Research suggests that food logging is a method that can be used to manage obesity [4]. Food logging consists of a person recording their food intake and recent methods use smartphones apps to make food logging process more convenient. Other methods use images for food logging, which can remove much of the complexity of food logging. The aim of this work was to investigate the feasibility of crowdsourcing of non-experts and experts in accurately determining calorie content in images of meals for food logging. The objectives were, 1) To determine if a crowd of experts and non-experts can accurately estimate the calorie content in images of meals, 2) To determine if analysing a group non-experts calorie estimations can be used to adjust calorie content in images of meals to promote accuracy. The remainder of this paper is as follows: related work will be discussed to determine what methods and technologies have been used previously, methodology section will be discussed detailing the processes in this work, a section detailing results of the experiments, discussion section examining the results, study limitations, and key conclusions.

2. Related Work

Crowdsourcing uses `wisdom of the crowds` to allow group of individuals to complete an activity to reach a goal or to solve a problem [12]. This technique has been applied to food logging in determining the calorie content and food quality of food in images [7]. Crowdsourcing was used to determine the food type, food size, and calorie content through using Amazon Mechanical Turk in [5]. In [5], tasks were repeatedly completed by workers using this platform to provide a nutritional workflow. Results from these experiments indicate that using crowdsourcing to determine nutritional value of meals is nearly as accurate as trained dieticians. Crowdsourcing was also used for dietary rating of food images [6]. In [6] a healthiness scale was used that allowed users to rate each image and results show there was a high correlation between user ratings and indicate that crowdsourcing can be used for dietary feedback. Similar research used a traffic light diet approach to assess the nutritional quality of images [7]. Results show that the ratings achieved high accuracy.
 (>75%) when examining all foods and that there is promise in utilising crowdsourcing for dietary feedback [7]. Other research used crowdsourcing to analyse menus of restaurant chain food menus to determine healthiness of food items [8]. Other related research also point to the use of crowdsourcing to provide novel methods for dietary management [9, 10]. Thus the literature indicates that crowdsourcing can be a useful technique for dietary management.

3. Methodology

The work described in this paper outlines two sets of experiments (1) compute descriptive statistical analysis (2) calorie adjustment process. The first experiments were completed to analyse the performance of non-experts and experts responses collected from a calorie estimation survey and determine the relationship between each group and meal images. Secondary experiments were completed that used the statistical metrics generated in the first experiments to adjust calorie estimations to enhance calorie accuracy. Accuracy results will be compared with ground truth calories of each meal image to measure performance of the adjustment process. For the calorie estimation survey, ethical approval was obtained by Ulster University.

3.1 Participants & Recruitment

Participants were invited to complete a calorie estimation survey. Participants were divided into two groups; experts and non-experts, experts were individuals who have knowledge of dietetics and nutrition, and non-experts individuals who have no trained knowledge of nutrition. Non-experts consisted of students within Ulster University and individuals not affiliated with Ulster University. Experts were recruited from the nutrition, and dietetic staff from Ulster University. Conveniences sampling was used (experts n=22 and non-experts n=120). Survey responses that were partially completed or participants that measured their food items in kilojoules (KJ) instead of calories were not included in this analysis.

3.2 Online Survey & Food Images

The online survey consisted of 15 photographs of meals taken by a researcher, who is also trained dietician. The 15 meals will include 5 breakfasts, 5 lunches, and 5 dinners, listed in Table 1. The photographs were taken on a smartphone device (iPhone 5). To calculate the calories of the food items in each image, each meal was weighed, and food labels and food tables were used. Participants completing the survey were asked the following question for each meal image: From viewing the photograph, enter the number of calories you consider is in this meal? Kcal OR KJ. To complete the survey, participants are asked to input their estimated calories for each meal image as well as confidence levels. In this work, calorie estimations were only used for analysis. Table 1 is a list of the meal images used in the online survey.

3.3 Preliminary Descriptive Statistical Analysis

Descriptive statistics were generated using calorie estimations for each group e.g. mean, mode, and median. Standard deviation for each meal image calorie mean for each groups was computed for each group to further highlight differences. Calorie differences were also calculated for each of the participant’s calorie estimation using the ground truth calories. Statistical analysis in this work was completed using Microsoft Excel version: 15.33.

3.5 Calorie Adjustment Statistics

Secondary experiments were completed that used descriptive statistics in the first experiments to adjust calorie estimations to enhance accuracy. To adjust the calorie estimates, the non-expert calorie dataset was used. A number of calorie statistics were first generated to adjust calorie estimates. The mean calorie estimate for each of the 15 meal images were generated. The overall calorie difference is also generated from all calorie estimates. The calorie difference is computed by subtracting the ground truth calorie from the calorie estimate, this will reveal a calorie difference or error. This was completed for each calorie estimate in the non-expert dataset. The following equations describe how each metric was computed.

\[
\text{Mean calorie for each meal image} = \bar{y} = \frac{\sum_{i=1}^{n} E_i}{n}
\]

(1)

\[
\text{Calorie difference} = D = C_{est} - C_{gt}
\]

(2)

\[
\text{Mean calorie difference} = \bar{x} = \frac{\sum_{i=1}^{n} D_i}{n}
\]

(3)

Equation (1) describes how the mean calorie is computed for each meal image type in each training fold, where \(E_i\) is a calorie estimate and \(y\) represents the mean calorie estimate for each meal image type and \(n\) is the number of estimations. Equation (2) describes how the calorie difference is calculated, where \(C_{est}\)
represents the calorie estimation and $C_{\text{gt}}$ is the calorie ground truth. Equation (3) is used to calculate the overall mean calorie difference using calorie estimations. Once $D$ has been calculated for each estimate using equation (2), $\bar{x}$ is calculated which represents the mean calorie difference. Figure 1 shows the calorie adjustment process using mean calorie difference, $\bar{x}$ and $y$. Figure 1 and 2 describe the overall process of calorie adjustment.

$$Error\ Percentage = \frac{C_{\text{gt}} - C_{\text{est}}}{C_{\text{gt}}} \times 100$$

(4)

where $C_{\text{gt}}$ is the ground truth calorie and $C_{\text{est}}$ represents the original mean calorie estimation for each meal, the percentage error for the adjusted meal calorie estimation was also calculated.

4. Experimental Results

4.1 Descriptive Statistical Results

Descriptive statistics were generated using experts and non-expert survey calorie estimations for meal images. The mean, mode, median, and standard deviation were computed to describe performance of each group in comparison to the ground truth. Figure 3 depicts the mean calorie estimations for each meal image for non-experts and experts group. The majority of mean calorie estimates for each meal image were greater than the ground truth for non-expert group, however there is a strong correlation for true calorie content and mean estimation for the majority of meal images for non-experts with a Pearson correlation of 0.88. This suggests that a crowd of non-experts are able to determine meals that have a higher calorie content than others.

Figure 4 compares the mean standard deviation for the non-expert group for each meal image with the ground truth calories. Analysis using non-expert participants suggest that the higher the ground truth calories, the higher the standard deviation. Pearson coefficient was calculated to describe this relationship using the mean calorie estimation for each meal image and the standard deviation for each meal and results show that the Pearson correlation coefficient was 0.80, which also indicates statistical significance.
Table 1 is a list of the food meals and calorie amounts. Table 2 lists descriptive statistics for both expert and non-expert groups for each meal image. T-tests were carried out to determine if there was statistical significance between the non-expert calorie estimations for each meal image and expert calorie estimations for each meal image, these results are listed in table 2.

Table 1: Meal Images and Calorie Amounts

<table>
<thead>
<tr>
<th>Meal Number</th>
<th>Meal Type</th>
<th>Calories (Kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bowl of cornflakes &amp; semi-skimmed milk</td>
<td>172</td>
</tr>
<tr>
<td>2</td>
<td>Breakfast fry-up</td>
<td>700</td>
</tr>
<tr>
<td>3</td>
<td>Branflakes with semi-skimmed milk &amp; banana</td>
<td>271</td>
</tr>
<tr>
<td>4</td>
<td>2 slices of white bread toast with butter &amp; marmalade</td>
<td>160</td>
</tr>
<tr>
<td>5</td>
<td>Hard boiled egg</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>Bowl of stew</td>
<td>400</td>
</tr>
<tr>
<td>7</td>
<td>Egg with mayonnaise &amp; watercress sandwich 2 slices wholemeal bread</td>
<td>315</td>
</tr>
<tr>
<td>8</td>
<td>Chicken, peppers, pesto basil leaves on white flat bread</td>
<td>295</td>
</tr>
<tr>
<td>9</td>
<td>McDonald’s Big Mac</td>
<td>495</td>
</tr>
<tr>
<td>10</td>
<td>Fish and chips</td>
<td>410</td>
</tr>
<tr>
<td>11</td>
<td>salmon, champ, carrots, peas &amp; sweetcorn &amp; white sauce</td>
<td>634</td>
</tr>
<tr>
<td>12</td>
<td>Chicken and pasta with sauce topped with cheese</td>
<td>400</td>
</tr>
<tr>
<td>13</td>
<td>Pork chop, champ and broccoli</td>
<td>349</td>
</tr>
<tr>
<td>14</td>
<td>Spaghetti bolognese</td>
<td>641</td>
</tr>
<tr>
<td>15</td>
<td>Chicken curry with white rice</td>
<td>577</td>
</tr>
</tbody>
</table>

Figure 3. Mean calorie estimation calculated for each meal image for each group (experts and non-experts) compared with ground truth.

Figure 4. Comparison of ground truth calorie for each meal and calorie standard deviation calculated using mean calorie estimation for each meal for non-expert group.

Table 2: Statistical Metrics Describing Experts and Non-Expert Estimations

<table>
<thead>
<tr>
<th>Meal Image</th>
<th>Non-expert mean (Kcal)</th>
<th>Non-expert mode (Kcal)</th>
<th>Non-expert median (Kcal)</th>
<th>Expert mean (Kcal)</th>
<th>Expert mode (Kcal)</th>
<th>Expert median (Kcal)</th>
<th>P-value (using expert &amp; non-expert estimations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>252.88 (±83.48)</td>
<td>200</td>
<td>250</td>
<td>205.91 (±48.71)</td>
<td>180</td>
<td>180</td>
<td>0.011</td>
</tr>
<tr>
<td>2</td>
<td>903.88 (±722.16)</td>
<td>800</td>
<td>800</td>
<td>719.32 (±32.67)</td>
<td>700</td>
<td>700</td>
<td>0.234</td>
</tr>
<tr>
<td>3</td>
<td>317.63 (±112)</td>
<td>300</td>
<td>300</td>
<td>273.41 (±38.34)</td>
<td>270</td>
<td>270</td>
<td>0.069</td>
</tr>
<tr>
<td>4</td>
<td>293.79 (±103.67)</td>
<td>350</td>
<td>300</td>
<td>191.59 (±48.93)</td>
<td>180</td>
<td>180</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>5</td>
<td>103.44 (±48.49)</td>
<td>90</td>
<td>90</td>
<td>80.23 (±1.88)</td>
<td>80</td>
<td>80</td>
<td>0.026</td>
</tr>
<tr>
<td>6</td>
<td>408.74 (±161.87)</td>
<td>500</td>
<td>400</td>
<td>394.23 (±34.44)</td>
<td>400</td>
<td>400</td>
<td>0.678</td>
</tr>
<tr>
<td>7</td>
<td>392.21 (±116.93)</td>
<td>400</td>
<td>400</td>
<td>377.73 (±69.50)</td>
<td>320</td>
<td>350</td>
<td>0.575</td>
</tr>
<tr>
<td>8</td>
<td>422.16 (±142.59)</td>
<td>400</td>
<td>400</td>
<td>302.95 (±22.40)</td>
<td>300</td>
<td>300</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>9</td>
<td>733.48 (±228.61)</td>
<td>600</td>
<td>700</td>
<td>553.86 (±127.47)</td>
<td>500</td>
<td>500</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>10</td>
<td>815.50 (±299.87)</td>
<td>600</td>
<td>765</td>
<td>517.73 (±133.34)</td>
<td>450</td>
<td>475</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>11</td>
<td>707.98 (±616.86)</td>
<td>600</td>
<td>650</td>
<td>601.36 (±32.63)</td>
<td>600</td>
<td>600</td>
<td>0.420</td>
</tr>
<tr>
<td>12</td>
<td>652.28 (±219.19)</td>
<td>600</td>
<td>600</td>
<td>412.50 (±26.65)</td>
<td>400</td>
<td>400</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>13</td>
<td>590.19 (±207.25)</td>
<td>500</td>
<td>550</td>
<td>418.64 (±72.61)</td>
<td>350</td>
<td>400</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>14</td>
<td>672.43 (±228.99)</td>
<td>800</td>
<td>650</td>
<td>637.73 (±50.89)</td>
<td>650</td>
<td>645</td>
<td>0.481</td>
</tr>
<tr>
<td>15</td>
<td>658.71 (±243.31)</td>
<td>550</td>
<td>600</td>
<td>602.36 (±56.07)</td>
<td>575</td>
<td>576</td>
<td>0.282</td>
</tr>
</tbody>
</table>
4.2 Calorie Adjustment Results

Five-fold cross validation was used to evaluate the calorie adjustment process. In these experiments, each fold was used as a testing dataset and the remaining were used as training to determine mean calorie differences and mean calorie estimations for each meal image. The mean calorie estimation calculated for each meal was used as a baseline, as outlined in Figure 1 and 2. Figure 5 shows the results of original mean calorie difference and adjusted mean calorie difference for each test fold. The adjusted mean calorie difference was calculated using equation (2) and (3) for each fold and compared with original mean calorie difference for the same fold. Each fold achieves a lower mean calorie difference in comparison to the original estimations using the calorie adjustment process and the results using the rule based system, outlined in figure 2, demonstrates that calorie accuracy improvement has been made. Figure 6 shows the comparison between the original mean calories, adjusted mean calories, and ground truth for each meal image. These results show that the calorie adjustment method increases the accuracy of user meal estimations.

![Figure 5. Original mean calorie difference compared to adjusted calorie difference for each test fold.](image)

Figure 5. Original mean calorie difference compared to adjusted calorie difference for each test fold.

![Figure 6. Comparison of mean adjusted calories against mean original calories along with ground truth calories for each meal image.](image)

Figure 6. Comparison of mean adjusted calories against mean original calories along with ground truth calories for each meal image.

![Figure 7. Graph showing results of calorie deduction between mean original calorie estimations and mean calorie adjusted estimations for each meal.](image)

Figure 7. Graph showing results of calorie deduction between mean original calorie estimations and mean calorie adjusted estimations for each meal.

5. Discussion

The aim of this work was to investigate the feasibility of utilizing crowdsourcing as a method to provide dietary management for food logging. The methods presented in this work suggest that crowdsourcing can be used to adjust calorie estimations to promote accurate food logging. In regards to preliminary analysis of the collected calorie estimations, the non-expert (n=120) mean calorie estimations greatly overestimated the ground truth calories and the experts (n=22) overall mean estimates closely aligned with that of the ground truth. Analysis also showed that experts achieved a high accuracy in determining calorie content of different meal images. This is evident in the descriptive statistics analysis (Table 2) when compared to the ground truth calories.
The expert group also achieved greater accuracy in comparison to the non-expert group. Expert group mean calorie estimations were consistently accurate as shown in figure 3 and there was less calorie variance in the expert estimations as a low standard deviation was reported (in comparison to the non-expert group) as shown in table 2. For the non-expert group, analysis shows that greater standard deviations were reported for meal images with a higher ground truth calorie e.g meal image 2, 9, 10, 11, 14, and 15 had the highest ground truth calories and also had the highest calorie standard deviations in the non-expert group with 722.16, 228.61, 299.87, 616.86, 228.99, and 243.31 calories respectively. Pearson correlation tests were completed on non-expert dataset using the ground truth calorie and the calorie standard deviations for each meal image. Result show a Pearson correlation of 0.80 which suggests that there is a correlation between the ground truth calorie and the standard deviation of the meal images. Correlation coefficient tests were also completed using mean calorie estimations and ground truth calories for non-experts and this resulted in a coefficient of 0.88, which also shows statistical significance.

Descriptive statistics were generated using both expert and non-expert calorie estimations and there was a strong correlation between the mean calorie estimations and ground truth calories for each meal image, as shown in figure 3. The crowd of non-experts could collectively determine what meals had greater number of calories. These results echo other research completed in [6,7] in that crowds of individuals are able to determine healthy and non-healthy meals. T-tests were completed using calorie estimations for each meal for non-expert group and expert group and results show that the p-values for meal image 4, 5, 8, 9, 10, 12, 13 were found to be significant with p<0.05 in comparing the expert mean calorie and non-expert calorie for each image. Secondary experiments were completed that used crowdsourced statistical calorie metrics with the aim of adjusting calorie estimations to increase accuracy. This process was described in figure 2. Five-fold cross validation was used to evaluate this process. Results from these experiments show that the calorie adjustment method can reduce calorie estimations using two variables computed from each training fold; (1) overall mean calorie difference and (2) mean calorie for each meal image. This rule based system, highlighted in figure 2, was able to reduce individual calorie estimates to be closer to ground truth calories.

Table 3 lists overall mean calorie estimations for each meal image for non-expert group along with the adjusted mean calorie estimates for the same meal images. Error percentages were calculated to describe the performance between the original mean calorie estimates and the adjustment mean calorie estimate for each meal and results show an overall mean error percentage reduction from 40.85% to 25.52% was achieved using the calorie adjustment method. Figure 7 highlights what meal images experienced the largest mean calorie reduction which was meal image 4 and 13. Results of these experiments are shown in figures 5, 6, and 7 and these results show that this calorie adjustment method has potential to improve calorie estimations.
across user food calorie predictions. In regards to error percentage evaluation, meal image 4 (2 slices of white bread toast with butter and marmalade) had the largest reduction when comparing the adjusted mean calorie estimation to the ground calorie truth. Meal image 7 (egg with mayonnaise and watercress sandwich 2 slices of wholemeal bread) had the lowest error percentage when comparing the adjusted mean calorie with the original mean calorie estimates. The results presented in this work suggest that crowdsourcing can be used to adjust calorie estimates to enhance accuracy in relation to food logging.

6. Limitations

In regards to limitations of this work, bias is a major issue when discussing statistical relationships between data and in order to reduce bias in this study in regards to the calorie adjustment experiments, the overall mean calorie difference was calculated for all estimations instead of using specific meal image type mean calorie differences. These experiments were designed to minimise the bias by using a unified mean calorie difference computed using all estimations in the training folds. If specific meal image calorie differences were used (i.e. calorie difference for meal image 1) for calorie reduction then the adjusted meal image would be bias towards that specific meal image type. Bias could be further reduced by partitioning a number of meal image types for training and the remaining for testing, e.g. calorie estimations for meal images 1-8, and the calorie difference could be calculated using images 1-8. The remaining images (meal images 9-15) could be allocated as a testing dataset to test the overall mean calorie difference.

In this work, a mean calorie threshold was computed for each meal image using estimations in training sets, the threshold acted as a baseline to determine if the estimation was above this threshold then the estimation would be adjusted using the overall mean calorie difference. However, in some individual calorie adjustment instances, the calorie deduction deducted calories beyond the ground truth calorie point. To mitigate this issue, other calorie baselines could be explored instead of the mean i.e. mode or median as in this work some individual adjusted estimates fell below the ground truth calorie.

7. Conclusion & Future Work

The work presented in this paper explores how crowdsourcing can be used determine calorie content in food images. The aim of this paper was to investigate the feasibility of using experts and non-experts to determine calorie content in meals and how ‘collective wisdom’ can be used to adjust calorie estimations, however more work needs done in this area to investigate how crowdsourcing can be fully utilised to promote dietary management. We have highlighted several areas that could be addressed in future work; (1) extend the survey to allow more participants to complete the survey to gather more estimations for analysis, (2) the meal images used in this work contain a number of collective food items in one image, future work would extend the type of images and include individual food items i.e. slice of bread, single pieces of fruit. Research has been completed in [11] that suggests that individuals are able to estimate calorie of smaller food meals with greater accuracy compared to larger meals. With this knowledge, future work would include single food item images and smaller meal images to allow users to estimate calorie content and for analysis, (3) future analysis will include methods to reduce bias between calorie differences of meal images, e.g. a number of meal images type will not be included in the training set analysis and they will be used for testing e.g. estimations for meal images 1-10 will be used for training and the estimations for meal image type 11-15 will be used to test the calorie adjustment process. Other statistical metrics for calorie deduction i.e. mode, or median would also used instead of using mean calorie difference. The research presented in this work highlight how crowdsourcing calorie predictions and measuring their accuracy has the potential to be used for calorie correction to adjust estimations for more accurate dietary management.

7. References