e & g: Ji Wang

a & b: Oliver

c: Joanna

d: ?

Realistic scenarios

All the users of our system will carry a small camera with them. When a user consumes the food (including breakfast, lunch, dinner and snacks), the camera will automatically take a picture of the food and upload it into the public AWS S3 bucket to store the picture. We estimate that in the realistic scenario, the user will upload 15 images per day (three meals and snacks). We also use this estimate in our simulated system. Our system will transfer the picture from the public AWS S3 bucket to our AWS S3 buckets. After the transfer is completed, our system will calculated the healthy score of the food based on the picture uploaded by the user previously. The score will be stored in a database. The users can have access to their scoring details by logging in with their user ID on the website. The users can track their total score on all the food they consumed so far and their score for each food item (including time) at the same time. UPMC can also access to this information and use it to determine insurance quote.

### *Performance Testing*

1. White space

We apply two tests in this process. The first test deals with the accuracy of the calculation. The second test deals with different food items in the same picture. The third test deals with the file’s name of the pictures uploaded.

Before the first test, we originally set the points for unhealthy food as 30, the points for fair as 55, and the points for healthy as 80. When we use these benchmark numbers to calculate the the score of the fried chicken (shown in appendix), we got 40.48 ((97.1% \* 30 + 97.1% \* 30 + 70.1% \* 55 + 70.1% \* 55)/(97.1% + 97.1% + 70.1% + 70.1%)) (the recognition result is shown in appendix), which indicates that fried chicken is fair. The major problem with this result is underestimating the the unhealthy and healthy components of the tags. So we lower the points of healthy food to 10, and raise the points of healthy food to 90. With the same recognition result but more extreme points, we got 28.87 ((97.1% \* 10 + 97.1% \* 10 + 70.1% \* 55 + 70.1% \* 55)/(97.1% + 97.1% + 70.1% + 70.1%)), which falls into the unhealthy category and consistent with our common sense. We also try more than 200 pictures to verify our criteria, which seems to be justified.

Another significant component of our score calculation is the confidence level of tags. In our calculation, we use the confidence level as our weight of points. This issue is extremely important when dealing with multiple food items in the the same picture. As shown in the Appendix, the plate contains three food items, and this picture got 98.32 ((97.1% \* 10 + 97.1% \* 10 + 70.1% \* 90 + 70.1% \* 55)/(97.1% + 97.1% + 70.1% + 70.1%)), which is consistent with the common sense. Although this picture has been mistakenly recognized as dessert, we avoid the problem by assigning weight based on confidence level. Desert has much less weight than other fruits, which leads to a more accurate result. So it is reasonable to calculate the score of different food items in one picture using confidence level because of fault tolerance.

Before the third test, we find a bug when uploading some pictures into the S3 bucket. An error message will occur when the pictures with whitespace among the file names are uploaded. After comparing every characteristics, the only difference between successful trial and failed trial is the file name. We fix our code to make sure removing the whitespace before uploading.

Cost Analysis



