Detail On The Dataset (Recipe Corpus) Used In This Work

0.1 Dataset

To create a recipe data-set, we developed a web-scraper for geniuskitchen.com (web 2018). Our final data-set comprises of 230,876 recipes. Each recipe was stored as a plain-text document consisting information on ingredients, instructions, cuisine, cooking-time, cooking-approach, cooking-equipment, preservation-technique, meal-type (e.g., dinner, side-dish), dish-type (e.g., curry, salad), context-tag (festival, time-of-day), taste-tag (e.g. sour, spicy), advanced-tag (e.g., kids-friendly, week-night) and nutrition information. The nutrition information included : weight per serving (gm), number of servings, calories (kCal), protein (g), carbohydrate (g), sugar (g), sodium (mg), fat (g), saturated fat (g), calories from fat (kCal), cholesterol (mg), and dietary-fibre (g).

WHO score	Total (percentage) Recipes n =230872	FSA score	Total (percentage) Recipes n =230872
0	88(.04)	4	40877(17.71)
1	2732(1.18)	5	32227(13.96)
2	15603(6.76)	6	44807(19.41)
3	63969(27.71)	7	40647(17.61)
4	123453(53.47)	8	30754(13.32)
5	22395(9.7)	9	24249(10.5)
6	2533(1.1)	10	13002(5.63)
7	99(0.04)	11	4009(1.74)
8	0(0)	12	300(0.13)

Table 1: Distributions of our recipe corpus over WHO and FSA healthScores.

To assess how healthy each recipe is, we applied two internationally recognised healthiness guidelines: the WHO guideline for macro-nutrient intake goals (web 2008) and the FSA traffic light system for labelling food (web 2009). We extended (Howard, Adams, and White 2012)'s approach and considered 8 major macro-nutrients (e.g., proteins, carbohydrates, sugars, sodium, fats, saturated fats, fibres and cholesterol) and their corresponding ranges to define WHO based healthScore for recipes. For each recipe, each macro-nutrient is assigned 1 if it fulfils corresponding WHO range and 0 otherwise, sum of the scores for each macro-nutrient, accumulated in a final range of 0 (no nutrient goals are fulfilled) to 8 (all nutrient goals are fulfilled).

The FSA score corresponds to four macro-nutrients (e.g., sugar, sodium, fat and saturated fat) and uses a traffic light food labelling scale : green (healthy), amber (moderate) and red (unhealthy). To gain a single metric, we followed (Sacks, Rayner, and Swinburn 2009)'s approach of assigning an integer value to each colour (green=1, amber=2 and red=3) and the sum of the scores for each macro-nutrient, accumulated in a final range of 4 (very healthy recipe) to 12 (very unhealthy recipe). Table 1 details the spread of our recipe corpus over both healhScoring system.

References

Howard, S.; Adams, J.; and White, M. 2012. Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *BMJ* 345.

Sacks, G.; Rayner, M.; and Swinburn, B. 2009. Impact of front-of-pack traffic-light nutrition labelling on consumer food purchases in the uk. *Health Promotion International* 24(4):344–352.

2008. Who nutient intake goals. Accessed : March 2018.

2009. Fsa "traffic light system for food labelling available. Accessed : March 2018.

2018. Geniuskitchen. Accessed : March 2018.