

Learning to Estimate Nutrition Facts from Food Descriptions

Hadi Amiri, PhD; Andrew L. Beam, PhD; Isaac S. Kohane, MD, PhD
Department of Biomedical Informatics
Harvard Medical School, Boston, MA 02115

Introduction.

Risk factors such as body weight, blood pressure, and blood cholesterol can help people make informed decisions about their health promotion efforts. Food choices are among the most effective of these efforts which can help preventing chronic diseases, such as heart disease, diabetes, stroke, and certain cancers^[6]. Since different foods provide different energy and nutrients, healthy eating requires monitoring the nutrients that we consume. In addition, food offers new perspectives on topical challenges in Natural Language Processing and Computer Vision centering around computational models to extract nutrition facts from food-relevant textual content^[1] or find representations that are robust to occlusion and deformation in processing of food images^[2]. In fact, there is an emerging literature investigating *Food Computing* which aims to acquire and analyze food data from disparate sources for recommending and monitoring food consumption as well as addressing food-related issues in medicine, biology, gastronomy, and agronomy^[7]. The availability of large-scale food datasets and the above recent advances in Food Computing can transform the way that individuals consume food. Established in the literature is the ability to match foods with databases that contain nutrition facts^[1]. These approaches are effective for foods that exists in databases, but lack the ability to deal with large amount of *new* foods that don't exist in such databases; as reported in^[5], the average number of new foods per year is slightly less than 20K. In addition, there is a dearth of evidence as to whether learning food *ingredients* can help more accurate estimation of nutrition facts. Therefore, the aim of our study is two fold: (1) to develop effective computational models that accurately estimate nutrition facts of any given food, and (2) to investigate if computational modeling of food ingredients can help better estimation of nutrition facts.

Data Description. The USDA branded food products database^[4] contains food description, nutrition facts, and ingredients for a large number of foods which are voluntarily supplied by food industry organizations to USDA. USDA standardizes the reported nutrition facts by calculating nutrient values per 100 grams from those values provided per serving. The dataset contains about 237K food items, 40 nutrition fact types, and 100K ingredient types respectively. It exhibits a power law distribution, $c \times \exp(-0.14x)$, as some nutrition facts match with only a small number of foods. Although the majority of foods in USDA dataset contain important nutrition facts, some food organizations provide no or partial information about their products. In addition, the dataset is updated on a *yearly* basis^[4] and therefore many new foods do not exist in the dataset. These challenges inspires our work to develop computational models to automatically estimate nutrition facts from food descriptions.

Method. We developed multiple regression approaches including least squares Linear Regression (Linear) with L2 regularization (Ridge). In addition, the state-of-the-art approach for learning nutrition facts of foods given their descriptions was reported in^[1]—a Convolutional Neural Network (CNN) which used word n-grams to match food items with USDA dataset to derive nutrition facts. We extended this approach through *joint* learning of nutrition facts and ingredients of given foods. In particular, as Figure 1 depicts, we developed a multi-task learning framework to enable joint learning of ingredients and nutrition facts given food descriptions. Note that nutrition facts were normalized scalars and learned separately (see Discussion), and ingredients of each food were represented by a vector of 0/1s with 1 indicating existence of a specific ingredient in the food. The shared layers were used to exploit commonalities

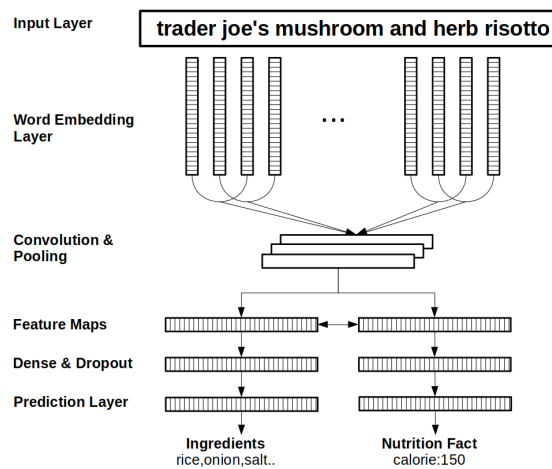


Figure 1: Proposed multi-task learning framework for joint learning of nutrition facts and ingredients.

and differences across tasks for more accurate learning. The network was trained by minimizing the following Mean Squared Error loss functions:

$$\mathcal{L}(I) = \mathcal{L}_{nutrition}(I) + \alpha \times \beta \times \mathcal{L}_{ingredient}(I) \quad (1)$$

where $\alpha \in [0, 1]$ controlled the extent to which ingredients contributed in overall learning of the task; $\alpha = 0$ indicates no contribution and was considered as a baseline here (basic CNN), and the parameter β was used to establish a common scale for loss magnitudes across tasks^[3]. Although, β could be tuned through grid search, we set $\beta = \mathcal{L}_{nutrition}^0 / \mathcal{L}_{ingredient}^0$ where \mathcal{L}^0 indicates loss at first iteration. Both loss functions optimized minimum squared error (MSE). Our framework enables learning semantic relations between food items and ingredients, e.g. learning that “roasted” foods should have “oil” as their ingredients, as well as semantic relations between foods and nutrition facts, e.g. learning that “rice” generally has high calories. Such relations are important indicators for accurate prediction of nutrition facts as they capture aspects of nutrients which may not be effectively represented in food descriptions.

Results. Our multi-task learning framework was trained and tested on each nutrition fact separately. For each nutrition fact, we partitioned food items into training data (80%), development data (10%) for parameter tuning, and test data (10%) for evaluation. We used grid search to optimize α for each nutrition fact using development data, then the resulting best model for each α was applied to the test data. Models were compared based on Coefficient of Determination (R^2 score $\in (-\infty, 1]$, where $R^2 = 1$ indicates perfect regression). Our multi-task learning framework outperformed Linear, Ridge, and basic CNN regressors on 70%, 55%, and 35% of nutrition fact categories respectively. In addition, Figure 2 shows the average R^2 performance over all nutrition fact categories for the top two best performing models, basic CNN and multi-task CNN respectively, across α values. As the results show, $\alpha = 0$ led to R^2 of 22.40, while there existed other α values, i.e. $\alpha \in \{.1, .4, .8, .9, 1.0\}$, that further improved the performance. We attribute this improvement to our model’s ability in utilizing semantic relations between food items, and their ingredients and nutrition facts.

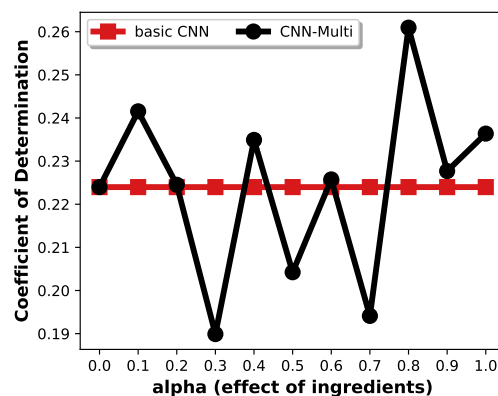


Figure 2: R^2 score regression performance; higher values are better.

Discussion. In this work, we developed an effective regressor to accurately estimate nutrition facts of foods. Our work highlighted the importance of learning ingredients for accurate estimation of nutrition facts. Our research has high value for developing diet monitoring applications, which may generate results with significant public health impact. Future investigations might explore associations among food *quantity* and *type* with ingredients and nutrition facts. In addition, ingredients often have a hierarchical form, e.g. *iodized salt*, *himalayan salt*, and *crystal salt* can all be mapped to the ingredient *salt*, which could be utilized to create a better semantic space for ingredients. In addition, our learning framework is trained on each nutrition fact separately; joint learning of these facts might create stronger regressors.

References

1. Korpusik, M., et al. 2017. Semantic mapping of natural language input to database entries via convolutional neural networks. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
2. Salvador, A., et al. 2017. Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).
3. Kendall, A., et al. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
4. US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. USDA Branded Food Products Database. Version Current: July 2018. Internet: <http://www.ars.usda.gov/nutrientdata>
5. New Food and Beverage Product Introductions of Consumer Packaged Goods. Retrieved on 3/12/2019. <https://www.ers.usda.gov/topics/food-markets-prices/processing-marketing/new-products/>
6. Eppinger, A., et al. 2018. Serum Marker of Glyphosate Exposure Associated with Changes in Oral and Gut Microbiome Composition. AMIA 2018 Annual Symposium.
7. Min W., et al. 2018. A Survey on Food Computing. arXiv. <https://arxiv.org/pdf/1808.07202.pdf>