



NUTRIOMICS CHALLENGE

Optimization of Dietary Nutrient Supplementation for Rational Rebalancing of Human Gut Microbiome

Zihan Wang, Ray Huang and Sang Yoon Byun

The Computational Problem

Select an optimal subset of nutrients for nutritional intervention that collectively promotes the growth of the maximal number of under-represented ASVs and the minimal number of in-range or over-represented ASVs. We'll discuss the general idea, then introduce our computational problem formulation.

We consider our test microbiome sample as a probability distribution of ASVs (since they are given in percentage). For each ASV, we know the region of normality that the percentage should fall into (see Section Subproblem). Then our core problem becomes a problem of finding the optimal subset of nutrients that makes the most ASVs fall into the normality ranges.

The given NIM table will decide how each ASV will be changed by a set of nutrients. We consider the changes in ASV values as the dot product of the NIs of the ASV with the nutritional intervention.

Sub-problem: $\mathbf{v}_i^{\text{low}}, \mathbf{v}_i^{\text{high}}$ are defined by a heuristic based on the given reference sample collection- a collection of ~ 1000 microbiome samples that are presumed to be mostly normal, comprised of ~ 2000 ASVs. One possible heuristic that we might use is for each ASV, we calculate the 10%, 90% quantile among the records, and set them as $\mathbf{v}_i^{\text{low}}, \mathbf{v}_i^{\text{high}}$. We consider these as already given, however, the heuristic may need to be further modified to be supported with biological knowledge. Note that the ASVs here could be a superset of the actual ASVs for an actual TMS.

Input: Number of ASVs, m , of around 300. Number of nutrients, n , of around 100. An NIM matrix $\mathbf{A} \in [0, 1]^{m \times n}$. A list of normal ASV ranges, $0 < \mathbf{v}_i^{\text{low}}, \mathbf{v}_i^{\text{high}} \leq 1, 1 \leq i \leq m$. A TMS $\mathbf{u} \in \mathbb{R}^+{}^m$ that is a (discrete) probability distribution over m items. The entries $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$.

Output: Consider a vector $\mathbf{r} \in \mathbb{R}^+{}^n, |\{i | \mathbf{r}_i \neq 0, 1 \leq i \leq n\}| \leq 10$, and let $\mathbf{b} = \mathbf{A} \times \mathbf{r}$. The failure counts c is defined as $c = m - |\{i | \mathbf{v}_i^{\text{low}} \leq \frac{\mathbf{u}_i + \mathbf{b}_i}{1 + \sum_{1 \leq j \leq m} \mathbf{b}_j} \leq \mathbf{v}_i^{\text{high}}, 1 \leq i \leq m\}|$. We want to find an \mathbf{r} that minimizes c .

Timeline

Week 4/5: Update problem formulation and work plan based on feedback received and begin software implementation. Week 6: Submit an extended description of the project and continue working on implementation. Week 7: Provide a summary of progress and preliminary results, and continue working on implementation. Week 9: Deadline for 5-page long paper.

Terms

Amplicon Sequences Variants (ASVs): A cluster of closely related microbial species that were collectively identified through 16s RNA sequencing. *Nutrient Impact* (NI): A value representing the anticipated capacity for a given microbe to benefit from supplementation of a specific nutrient. *Nutrient Impact Matrix* (NIM): A table that captures all the computed NI values for N nutrients are given over the entire set of ASVs. *Test Microbiome Sample* (TMS): A table representing the taxonomic profile of a subject with a set of deviant microbial species.