Brief Report for Flavor Mapping Analysis of Ingredients

Author: SC Wang

Summary

This study aims to demonstrate the potential of flavor mapping technique in flavor reformulation process. A total of 100 flavor ingredients with their notes information were collected from Good Scents Company Information System website. After data cleaning, normalized point-wise mutual information (NPMI) was calculated for every pair of ingredients according to their similarity in flavor notes. NPMI scores were further applied to construct flavor map. Flavor ingredients are clustered in the map based on their similarity in flavor profile. A total of 6 clusters are identified, which corresponds to 4 major flavor families (Roasted; Sweet/Fruity/Floral; Mint; Dairy). This flavor map could play a significant role in flavor reformulation process to guide the choice of flavor alternatives.

Methods

This study was completed by using Python with Spyder IDE. A total of 100 commonly used flavor ingredients were applied in this study. Their flavor notes were collected from "The Good Scents Company Information System" (website: here). A subset of 20 flavor ingredients is provided for demonstration (refer to data folder).

A series of data cleaning steps were performed for flavor notes sentences (*Variable: Flavor Notes/notes*). After tokenization, stop words were removed. Stop word list from NLTK was extended to include the following words: *'slight', "nuances", "nuances.", "note", "notes", "nuance", "slightly", "undernotes", "back", "background", "de", "impact", "like", "mild", "strong"*. Lemmatization was then applied to remove inflectional endings of flavor notes by using WordNet Lemmatizer. However, lemmatization failed to clean all words with similar meanings due to typo or other issues, such as "fruitti", "fruit" vs. "fruity"; "jasmin" vs. "jasmine". All of these words were standardized to the corresponding common notes (say "fruity", "jasmine"). Finally, the flavor notes were converted to bag-of-words with individual words (flavor notes) as columns and values corresponding to frequency of occurrence in each ingredient (0 or 1, Figure 1).

	acetic	acidic	aldehydic	almond	...	waxy	wine	woody	name
0	0	0	0	0	...	0	0	0	2-methylpyridine
1	0	0	0	0	...	0	0	0	2-pentenal
2	0	1	0	0	...	0	0	0	acetic acid
3	0	0	0	1	...	0	0	1	benzaldehyde
4	0	0	0	0	...	0	0	0	benzyl alcohol
5	1	0	0	0	...	0	0	0	butyric acid
6	0	0	0	0	...	0	0	1	caryophyllene
7	0	0	0	0	...	0	0	0	citral
8	0	0	0	0	...	0	0	0	citronellol
9	0	0	0	0	...	0	0	0	coumarin
10	0	0	0	0	...	0	0	0	d-limonene

Figure 1. An example of cleaned dataset

Exploratory analyses were conducted to find the common flavor notes in the dataset and the flavor ingredients with rich flavor profile. To obtain flavor mapping, normalized point-wise mutual information (NPMI) was computed for every pair of flavor ingredients in the dataset based on how frequently flavor notes appear in both ingredients. NPMI ranges from -1 to 1 with -1 for never occurring together, 0 for independent, and 1 for complete co-occurrence. Ingredient pairs with NPMI of 0.55 and above were used to generate flavor map. After initial exploration, 6 major clusters were visually identified from the map and labelled in the dataset accordingly to generate the final flavor map.

Results

Figure 2 shows the most frequent flavor notes in the dataset. Over 35% of the common notes belong to fruity flavor family: "fruity" has the highest frequency; "citrus", "apple", "berry" and "pineapple" notes also fall in this category. "Sweet" and "green" notes also occur more frequently than others. Some ingredients have more flavor notes than others as shown in Figure 3. Trimethyl pyrazine, methyl pyrazine and alpha-ionone have the richest flavor profile with 12 unique flavor notes.

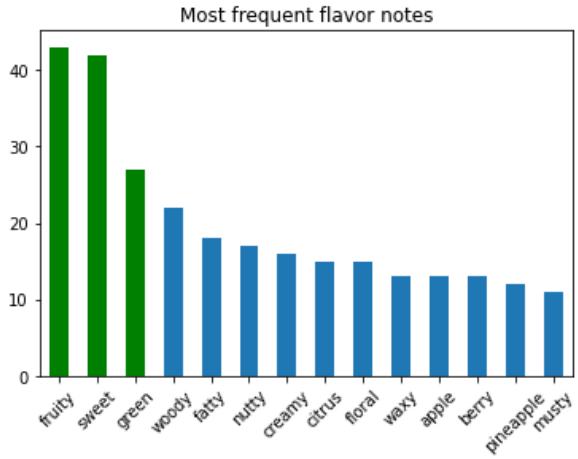

Figure 2. Most frequent flavor notes among 100 ingredients

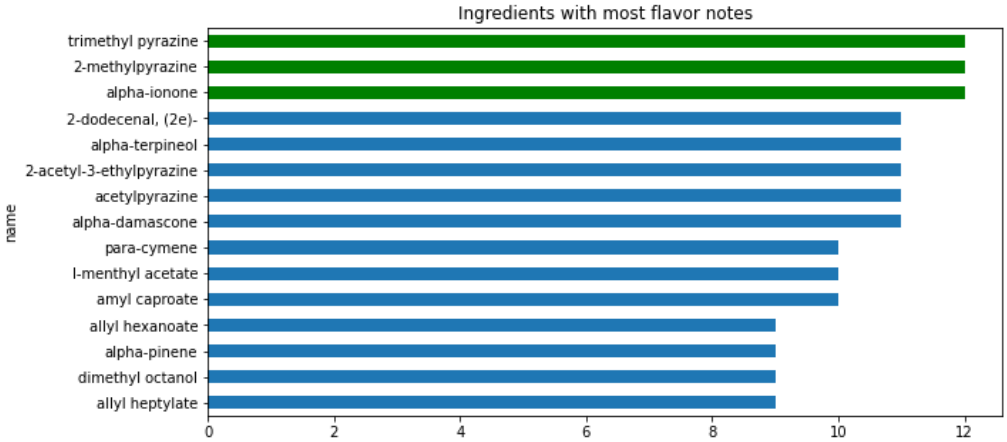

Figure 3. Ingredients with most flavor notes.

Flavor map is visualized in Figure 4. If two ingredients are more similar to each other in their flavor notes, they appear closer in the map. According to the distribution, 6 clusters of ingredients are noticed. Group 1 (Figure 4a, 4b) have 9 ingredients with roasted flavor notes. They all belong to pyrazine or pyridine chemical category, which are commonly present in roasted or baked products (such as coffee, roasted nuts, bread etc.). Group 2 (Figure 4a, 4c) contains majority of the ingredients. Although these chemicals are from different categories, they all share the notes of fruity, sweet, and/or floral. Group 3 and 4 (Figure 4a, 4b) have 5 ingredients in total. They are related to mint, providing green flavor and/or cooling sensations. Group 5 & 6 (Figure 4a, 4d) consist of 22 ingredients with flavor notes of creamy, fatty, dairy, milky. They are naturally present or commonly added in dairy products and products with dairy flavor.

This flavor map could be very useful in guiding the selection of ingredient alternatives from the flavor perspective. For example, due to some restrictions, a company has decided to remove veratryl aldehyde from all the products. However, veratryl aldehyde (Figure 4d, underscored) is a significant ingredient for sweet, dairy, creamy, and vanilla flavor notes. To compensate for potential flavor loss result from this removal, it is better to find some substitutes with similar notes. Ethyl vanillin, vanillin, anisaldehyde, and nonanoic acid are potential candidates according to their proximity to veratryl aldehyde in the flavor map. Further experiments could be conducted to select the best possible alternative. Overall, this flavor mapping approach would largely save the time and effort required in the flavor reformulation process.

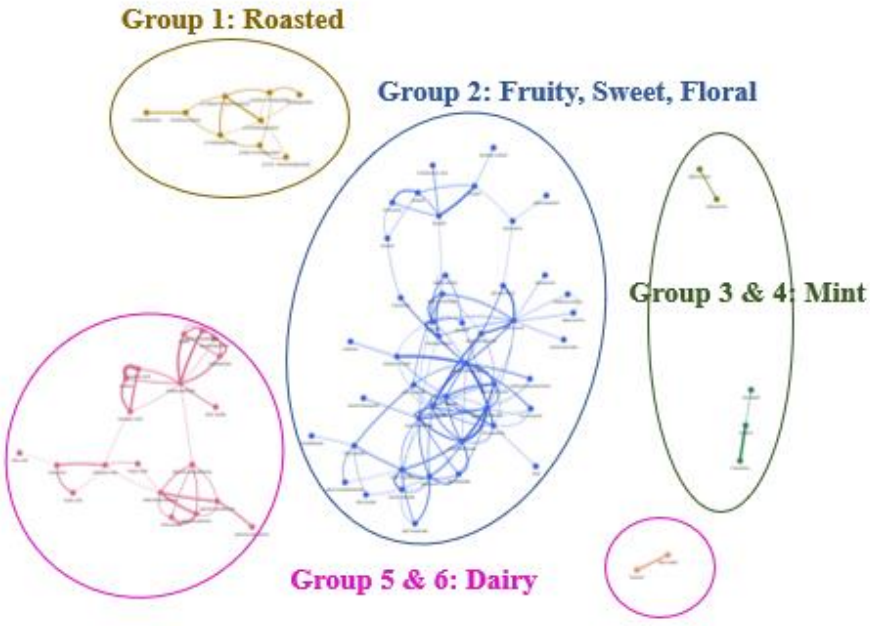

Figure 4a. Flavor mapping for ingredients with NPMI ≥ 0.55

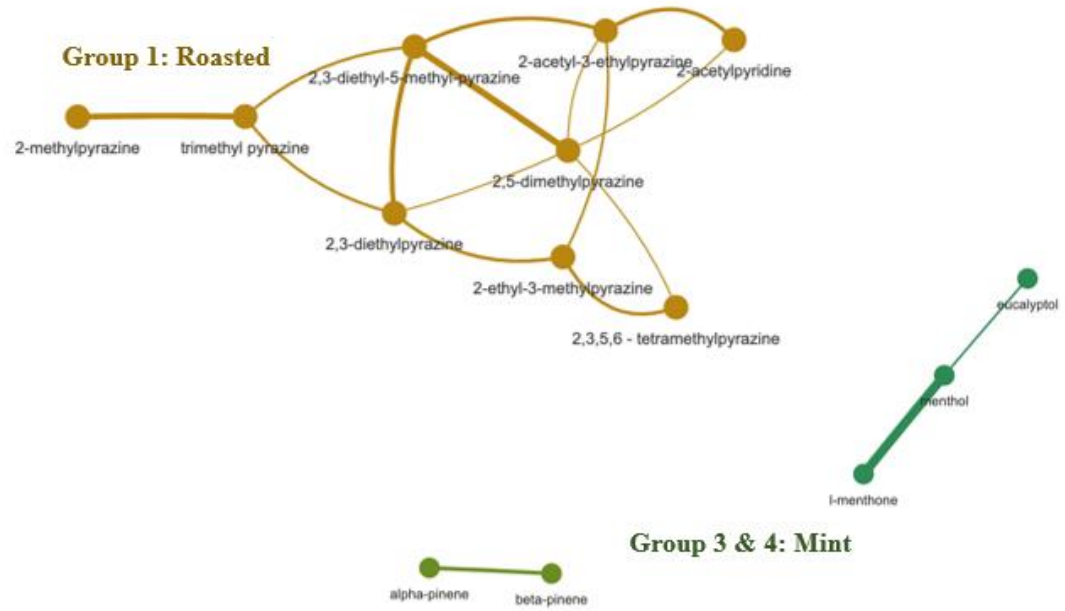

Figure 4b. Group 1, 3, and 4 in flavor map

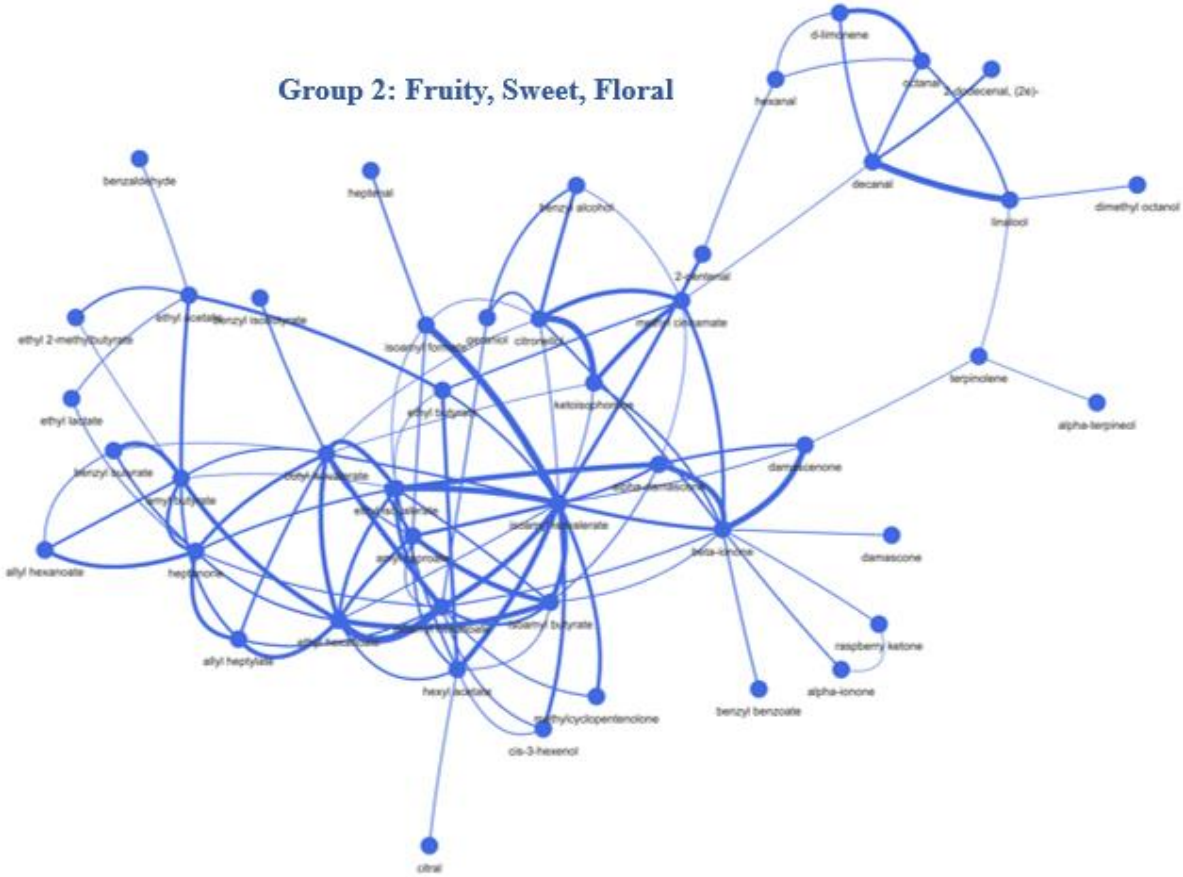

Figure 4c. Group 2 in flavor map

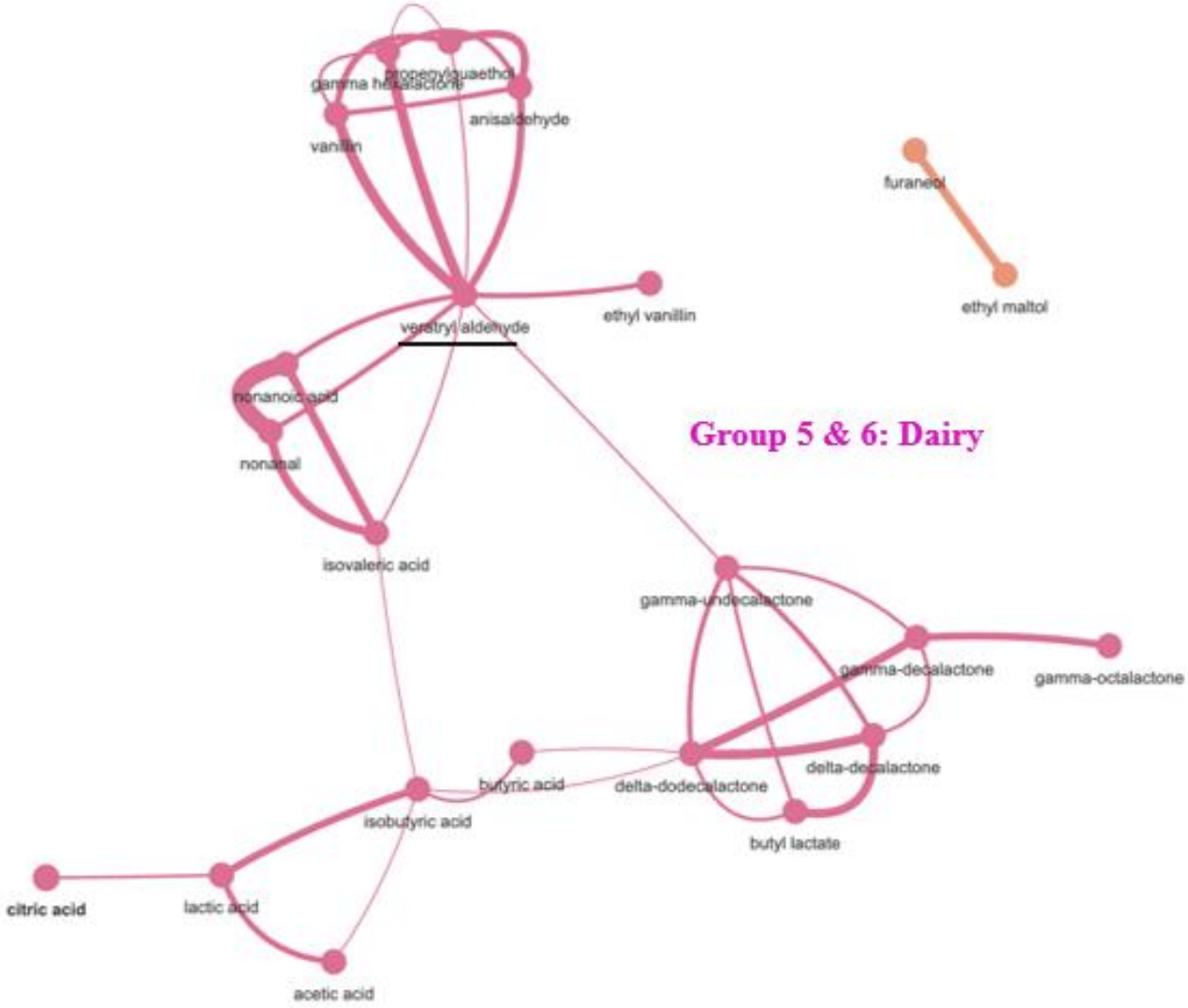

Figure 4d. Group 5 and 6 in flavor map