W51.P2

Bag-of-Foods: Analysis of Personal Foodlogging Data

Yuji Goda† Sosuke Amano† Yoko Yamakata† Kiyoharu Aizawa†

†Dept. of Information and Communication Eng., The University of Tokyo

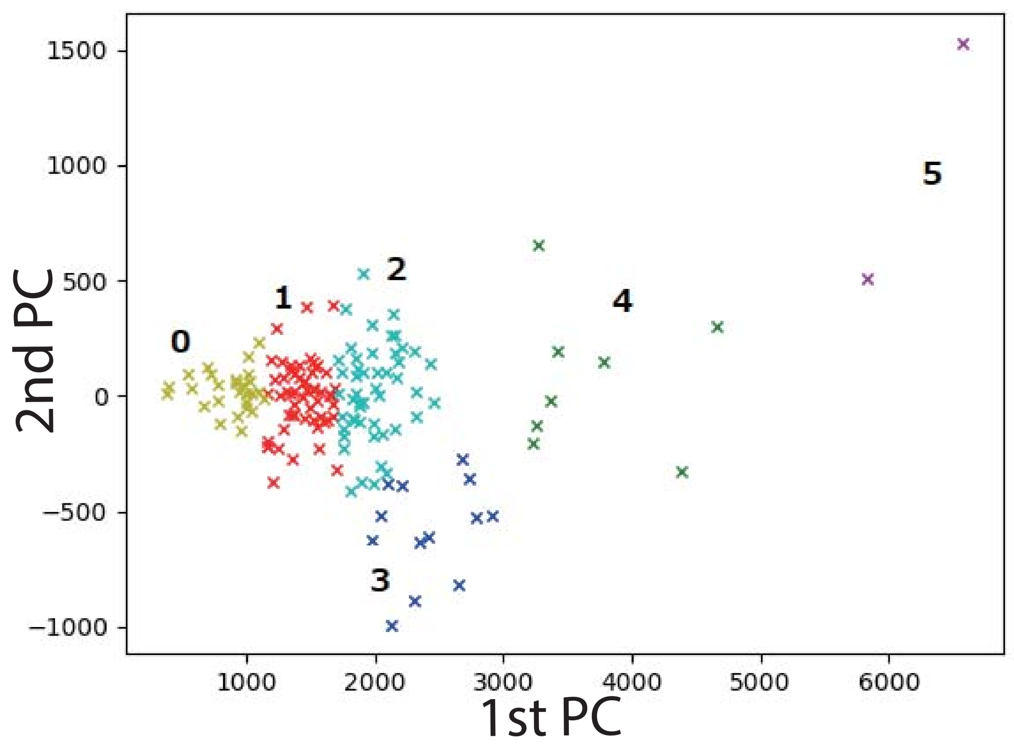
Motivation

- Foodlogging tool, FoodLog, has collected records more than 1.5M records
- We have proposed Bag-of-Foods (BoF) to represent ones' diet preference using users' record rather than Food Frequency Questionnaire

Nutrition-based User Clustering

- Each meals represented by 31 nutrition values
  - Calorie, fats, carbons, proteins, minerals, vitamins
- 157 users clustered to 6 groups by k-means clustering

PCA result



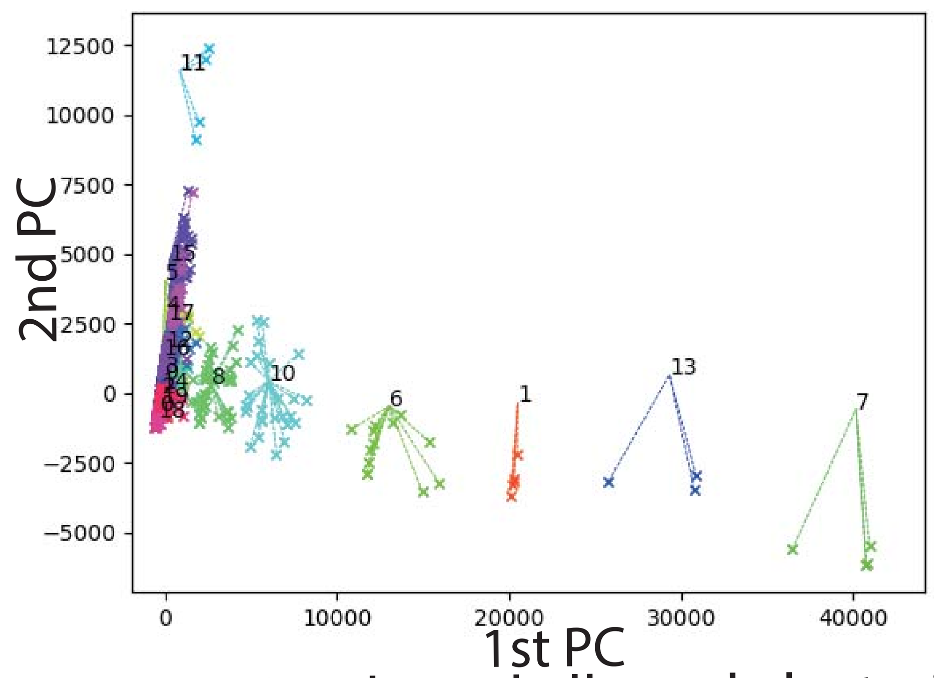
Cluster 0 and 4+5 show significant differences according to the actual meals listed in the table

Typical example of meals users in each cluster eat

Cluster	Meal	Cluster	Meal	Cluster	Meal
0	Jam & bread, Coffee	2	Egg soup, Fried rice	4	Curry & Rice, Fresh salad, Milk
	Rice cake		Yogurt, Coffee, Sandwich, Vegetable juice		Fried vegetables, Rice, Grilled fish, Miso soup
1	Bread, Fried egg, Tomato	3	Banana, Coffee, Cup of fruits, Yogurt, Fresh salad, Jam & bread	5	Rice, Miso soup, Fried chicken, Stewed kelp, Grilled fish, Cut cabbage
	Chinese soup, Rice, Natto (soy beans)		Scone, Quiche		Fried shrimp, Cut cabbage, Boiled egg, Rice, Doughnut, Rice vermicelli, Miso soup, Boiled spinach, Chili shrimp

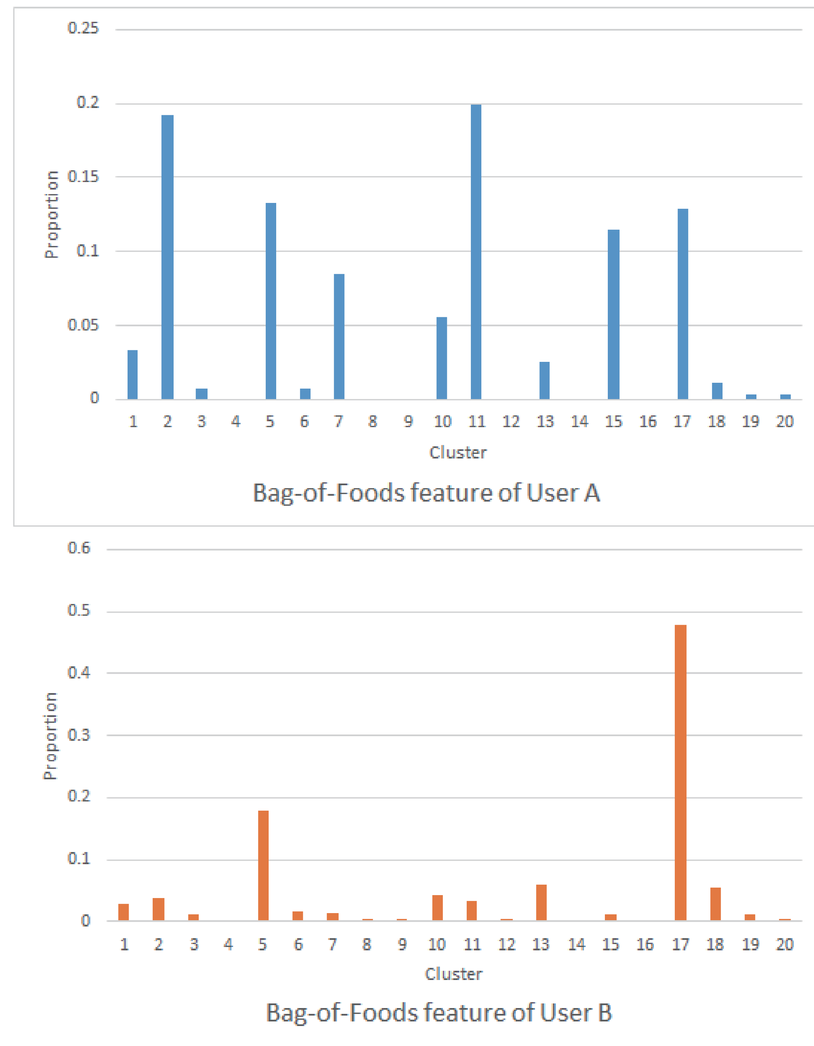
What's Bag-of-Foods

- Previous results shown nutrition-based clustering give users' preference
- Vectorize users' diet preference with Bag-of-Words method
  1. All meals clustered by k-means based on 31 nutrition value
  2. Each user is characterized with the frequency of meals of each cluster the user intaked



nutrition-based all meal clustering

Clustering gave interpretable "words" of meals



Examples of Bag-of-Foods feature of 2 users when # clusters = 20

Evaluation of Bag-of-Foods

Is Bag-of-Foods able to calculate similarity of users' diet preferences?

Generated the data of similar & dissimilar users for evaluations as follows

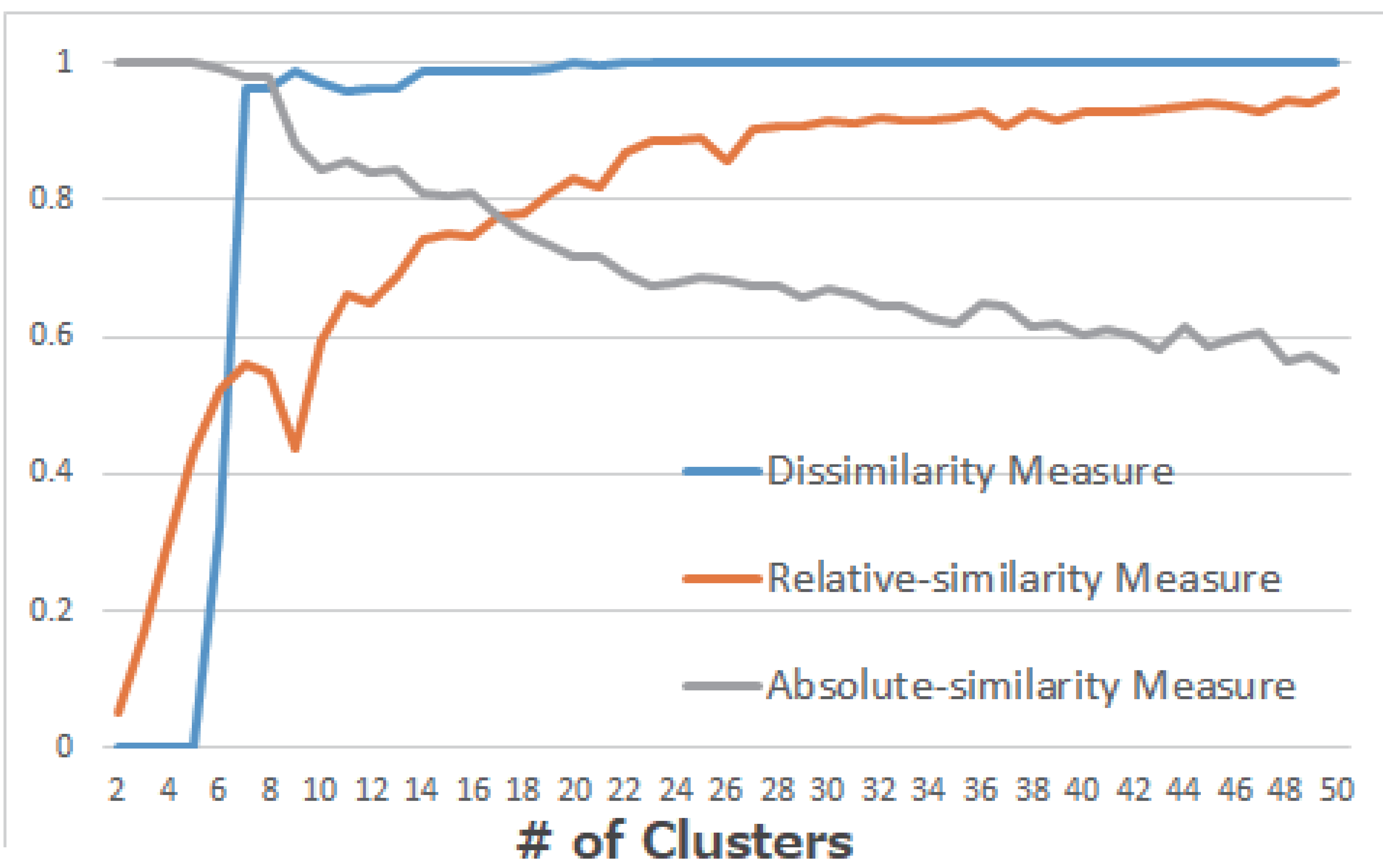
Similar user:
Devide one uses' foodlogging data into 2 virtual users

Dissimilar user:
The users specified as dissimilar users innutrition-based user clustering above (Cluster 0 vs. Cluster 4 + 5)

Measures for evaluation

Dissimilarity measure:
Proportion of dissimilar users placed far in feature space

Relative-similarity measure:
Proportion of users whose similar user' s Bag-of-Foods is closer than any other users



3 measurement scores in each number of clusters

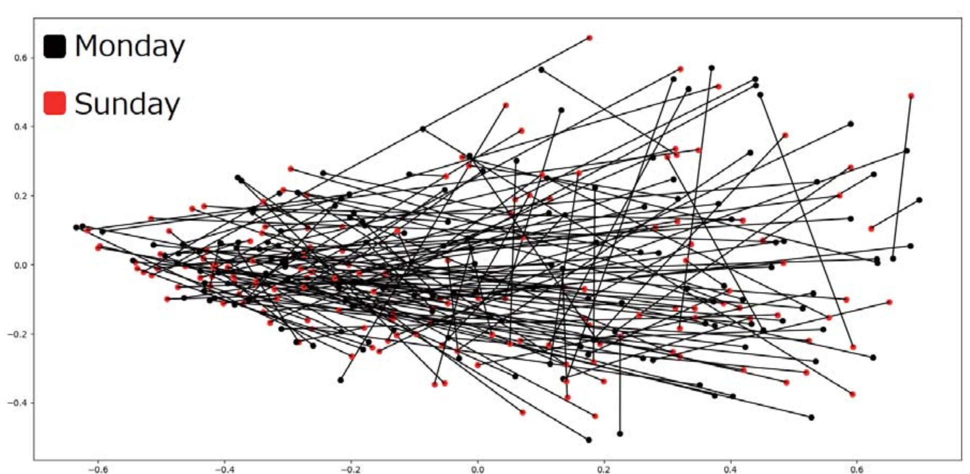
How Different?

	User A	Similar User	Dissimilar User
Meals in sequence	Banana	Dry mango	Egg Sandwich
	Green salad, Rice boll	Cereals, Banana, Yogurt	Ohagi
	Nyumen noodles, Sweet beans, White raddish stew	Tomato spaghetti, Fried chicken	Baked cheese cake, Custard pudding, Black coffee
	Fried vegetables, Tuna sashimi	Fried vegetables, Tuna salad, Miso soup	Pancake, Black coffee
	Banana, Grape	Yogurt, Banana, Cereals	Punpkin stew, Miso soup, Grilled fish, Rice

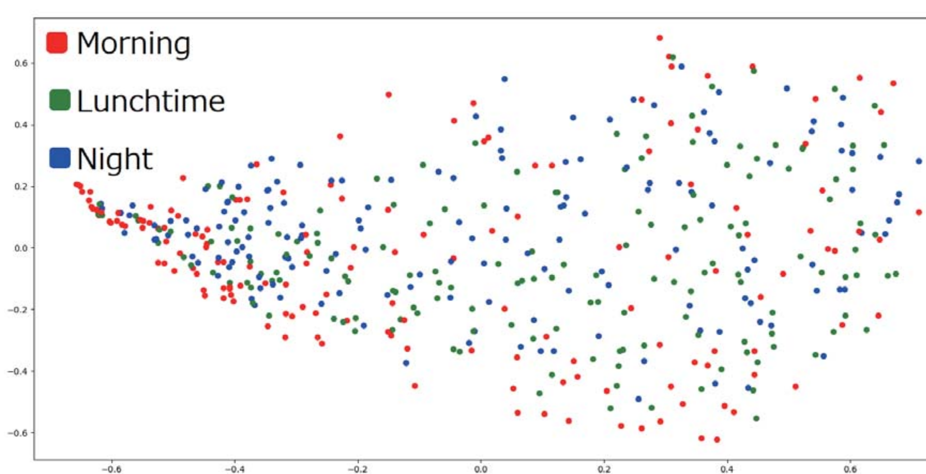
3 users' actual foodlogging data in sequence

Diet Preference Analysis using Bag-of-Foods

Users' foodlogging data devided by each timescale to visualize time-specific behavior



Scale: day in the week
BoF feature show huge daily diversity within a user



Scale: time in a day
Morning meals are deviated compared to other times

Conclusion

BoF with 20 words found 83.1% of similar user
→ BoF is effective for personal diet preferece representation
Even for the same users BoFs change drastically on the differencet days

Future Work

Extend BoF to Temporal (monthly, weekly, daily) BoF to analyze temporal variation of users' diet preference