Personalized Food Recommendation as Constrained Question Answering over a Large-scale Food Knowledge Graph

Yu Chen¹, Ananya Subburathinam¹, Ching-Hua Chen², Mohammed J. Zaki¹

 1 Rensselaer Polytechnic Institute, Troy, NY, 2 IBM Research, Yorktown Heights, NY

Contributions

- KBQA-based personalized food recommendation framework.
- Personalization by automatically applying persona focused nutritional guidelines and ingredient constraints.
- Novel techniques to effectively handle numerical comparisons and negations

Template-based Natural Language Question						
Generation						

- Results
- **Goal**: Realistic benchmark questions that reflect real word requirements and constraints.
- Generated templates based on analysis of submissions related to recipe and diabetes made on the social media forum, Reddit (http://www.reddit.com/).
- Example: The following image shows an annotated user query from
- Our PFOODREQ model is tested against our main baseline BAMnet [3], an embedding-based KBQA method MatchNN motivated by [4], A Bag-of-Word (BOW) [5] vectors based KBQA baseline as well as their personalized versions.

- in the queries.
- A QA style benchmark for personalized food recommendation based on a large-scale food KG and health guidelines.

Personalized Food Recommendation

- Factors influencing food recommendation:
- Users' explicit requirements
- Crucial health factors (e.g., allergies and nutrition needs
- Rich food knowledge for recommending healthy recipes
- User's dietary preferences and health guidelines

Personalized KBQA Benchmark

• A benchmark QA dataset based on the extensive *FoodKG* [1] knowledge graph.



- Based on the analysis, we identify the four common types of constraints:
- positive ingredient constraints stating what ingredient(s) can be included in the recipe,
 negated ingredient constraints stating what ingredient(s) cannot be included,
 nutrient based constraints such as "low carb" or "high protein", and
 cuisine based constraints such as "Indian", or "Mediterranean".
- Queries feature a combination of these constraint types.
- **Example Templates** and the queries generated from them are shown in the table below.

Templates

- **1.** What are {tag} recipes that contain {in_list}?
- **2.** What $\{tag\}$ recipes can I cook without $\{in_list\}$?
- **3.** Recommend {limit} {nutrient} {tag} recipes which have {in_list}?

Generated Questions

- **1.** What are *jellies* recipes that contain *orange*?
- **2.** What *turkish or dinner-party* recipes can I cook without *canned milk*?
- **3.** Recommend *low protein russian* recipes which have *onions*?

Method	MAP	MAR	F1
BOW	2.1	2.0	2.3
MatchNN	2.7	2.7	3.0
BAMnet	3.1	3.0	3.1
P-BOW	4.5	4.4	4.2
P-MatchNN	45.5	45.1	41.2
pfoodReQ	62.7	61.8	63.7

• Ablation study to investigate the impact of different model components for our PFOODREQ method demonstrates the importance of incorporating rich personal information.

Method	MAP	MAR	F1
PFOODREQ	62.7	61.8	63.7
– KA	58.5	57.8	58.6
$- \mathrm{CM}$	29.7	29.3	25.9
- QE	4.5	4.5	4.2

- We also support the concept of personalization based on a user
 food-log by leveraging recipe embeddings [6].
- Similar recipes are identified through a recipe's nearest neighbors in the embedding space.
- KG Subgraph Expansion to include these similar recipes as candidate answers

- Each example in the dataset contains a *user query, dietary preferences, health guidelines* associated with the user, and the *ground-truth answers* (i.e., recipe recommendations).
- Ground-truth answers are those recipes from FoodKG that satisfy both explicit requirements and personalized requirements.
- Template-based Natural Language Question Generation: with explicit requirements.



Populated questions: What {Indian} dishes can I make with {chicken}?

• Health Guidelines: Carefully select some food-related guidelines from ADA

• We split the personalized KBQA benchmark dataset into training (4,621 examples), development (1,540 examples) and test sets (2,269 examples).

Model

- A personalized food recommendation system is supposed to take as input a natural language question, dietary preferences as well as health guidelines, and retrieve all recipes from a food KG that satisfy the requirements contained in the input.
- \bullet The overall architecture of our $\mathbf{pFoodReQ}$ framework is shown below:



- Answer Ranking prefers those recipes that not only satisfy various query constraints, but also are similar to the user's food log.
- Experimental results to demonstrate benefits of log based similarity.

Method	MAP	MAR	F1
pFoodReQ	27.3	26.4	32.6
PFOODREQ +RecipeSim	34.5	33.0	36.6

References

- [1] Steven Haussmann, Oshani Seneviratne,
 Yu Chen, Yarden Ne'eman, James Codella,
 Ching-Hua Chen, Deborah L McGuinness, and
 Mohammed J Zaki.
- Foodkg: A semantics-driven knowledge graph for food recommendation.
- In International Semantic Web Conference, pages 146–162. Springer, 2019.
- [2] American Diabetes Association.
 - 5. lifestyle management: standards of medical care in diabetes-2019.

- lifestyle guidelines [2] pertaining to nutrient budgets.
- Convert each selected guideline to structured representation, as shown in the example below:



• Dietary Preferences: Randomly generate dietary preferences on ingredient likes and allergies based on food tag in each query.

- **pFoodReQ** architecture has four important modules: Query Expansion (QE), KG Augmentation (KA), Constraint Modeling (CM), and KBQA modules.
- Query Expansion: This module appends personal constraints to raw user query.
- KG Augmentation: Dynamically augments a KG subgraph based on the results of symbolic number comparison before feeding it to a neural network-based KBQA system.
- **Constraint Modeling:** Explicitly indicate the existence of a negative constraint to a KBQA system, generate and concatenate constraint embedding to the word embeddings.
- **KBQA Module:** Can use any KBQA method that accepts a query and a KG subgraph, and returns the most relevant answers from the KG subgraph.
- The main KBQA module we use in this work, is a neural network model, based on the Bidirectional Attentive Memory Network model (or BAMnet) [3]

Diabetes care, 42(Supplement 1):S46–S60, 2019.

- [3] Yu Chen, Lingfei Wu, and Mohammed J Zaki.
 Bidirectional attentive memory networks for question answering over knowledge bases.
 In NAACL, 2019.
- [4] Antoine Bordes, Sumit Chopra, and Jason Weston.

Question answering with subgraph embeddings.

In *EMNLP*, pages 615—-620, 2014.

- [5] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge university press, 2008.
 [6] Diya Li and Mohammed J Zaki.
- [6] Diya Li and Mohammed J Zaki. RECIPTOR: An effective pretrained model for reciperepresentation learning. In *SIGKDD Conference*, 2020.

*This work is supported by IBM Research AI through the AI Horizons Network