Embedding the World of Food in Our Minds

: Constructing a Food Schema

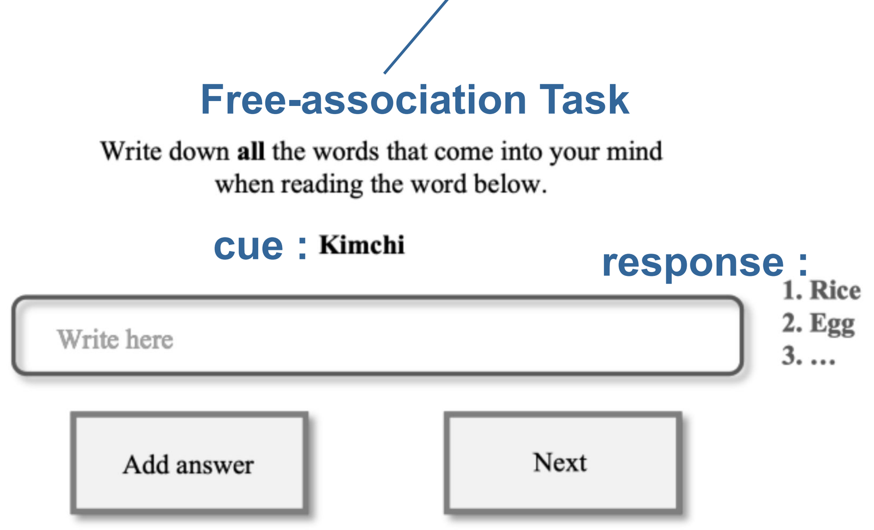
Hyunsu Jung, and Inhwa Oh, School of Data Science, Hanyang University
Hyungwook Yim

CCD Computational Cognition & Development Lab

Research Background

Why Do we Need to study Food Schema?

- People hold different food schema for the same food, and studies suggest that these differences may relate to eating habits and weight management.



Prior Work

- Responses were manually labeled into five semantic categories.
- Cue one-hot vectors + category proportions were used to train a normal vs. obese classifier.

→ Indicates group-level differences in food schema.

Limitation of Prior Work

- **Lack of Domain-Specific Embeddings**
  Korean food-specific terms (e.g., *pajeorim*, *chonggak-kimchi*) were not covered by existing embeddings.
- **Manual Category Tagging**
  Labor-intensive and not scalable.

Our Approach & Why It Matters

Food Domain corpus construction

- Food-related blog texts : contain rich relationships between foods & daily contexts (places, sensations, situations).

Word Embedding

- Train embeddings for Korean food/brand/place that were not covered in general models.

Automatic Category Prediction

- Propose an automated tagging pipeline suitable for large datasets.

Embedding based Normal/Obese Group Prediction

- Pilot test of whether embedding-based representations yield improved predictive performance.

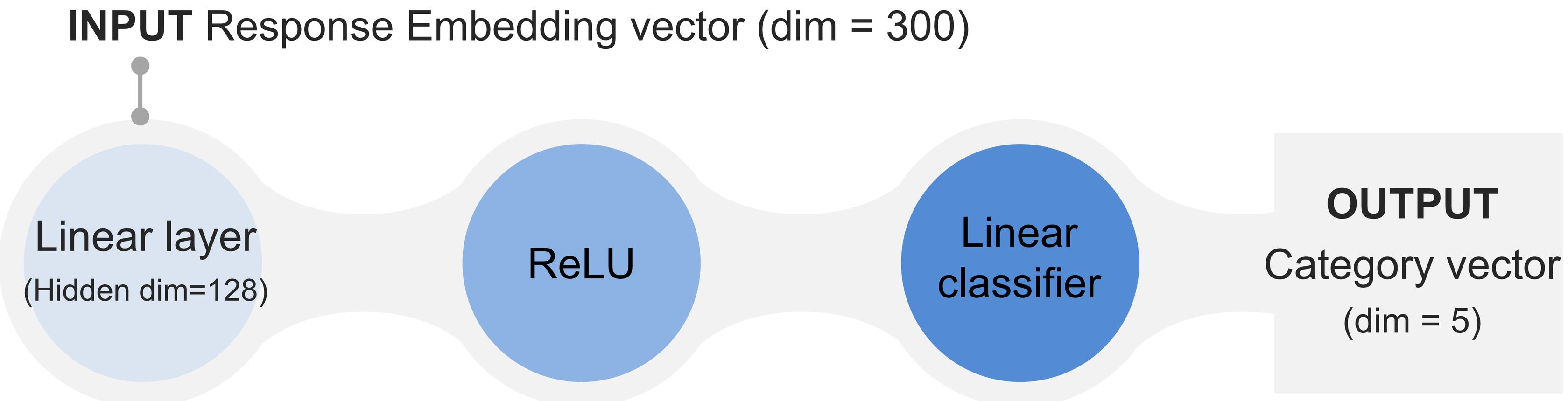
Methods

Corpus & Word Embedding

- **Corpus Construction**
  Crawling Food-Related Blog Texts
  Search keyword : 135 cue, 1,632 response / Total documents : 424,080
- **Embedding Training**
  Word2Vec
  Captures semantic relations in food-related contexts
  Output: 300-dimensional vectors
- **Embedding Quality Check**
  Hierarchical Clustering
  MDS Visualization
  Similar Words

Category Predict Model

INPUT Response Embedding vector (dim = 300)

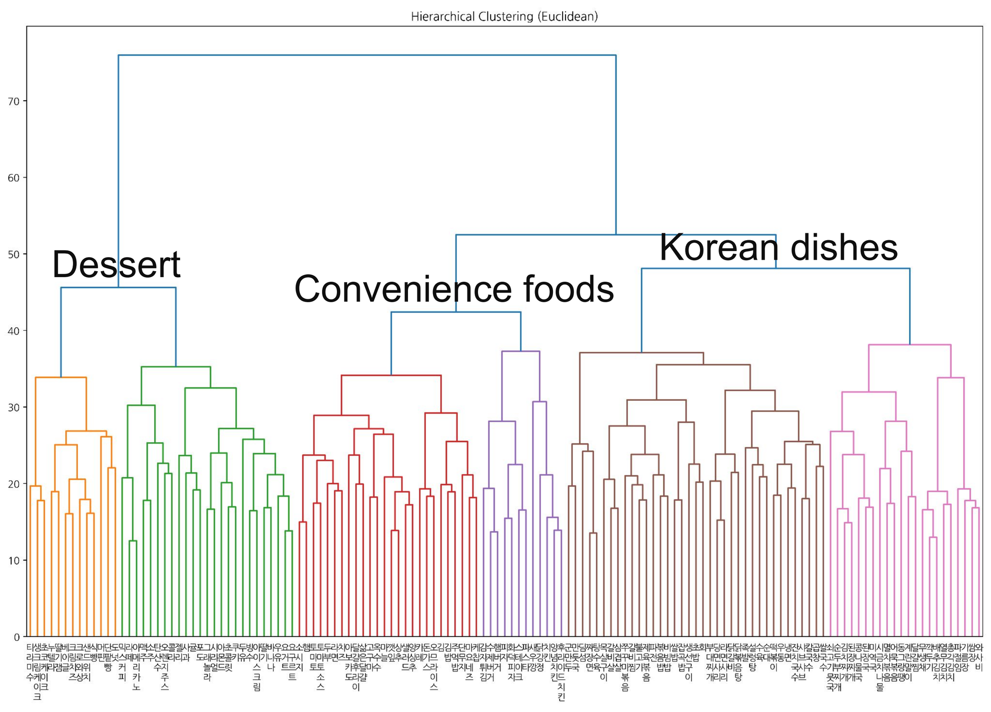


Normal/Obese Group Predict Model

INPUT : Cue+Response Embedding vector (sequence length = 2)
Model Architecture : RNN (300-128-1)
OUTPUT : normal(0) / obese(1) group

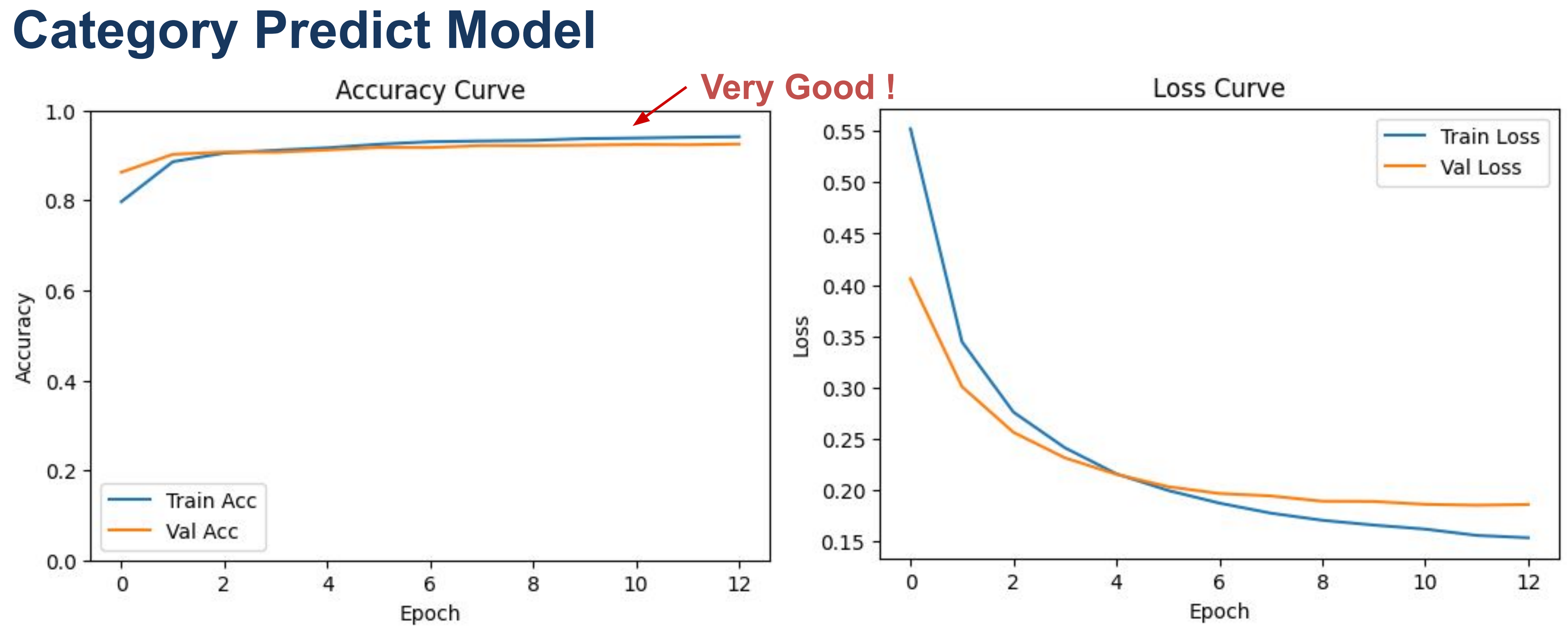
Results

Corpus Construction



- **Hierarchical clustering food cue embeddings**
  Meaningful groupings emerged
- **MDS projection of Cue word**
  Cue words formed clusters with semantically related terms (foods, places, store names, etc.).

Category Predict Model



- **MLP Category Classifier**
  e.g
  Input: embedding vector of *"rice"* (300-dim)
  Output: 10000 (5dim) = food word category

Normal/Obese Group Predict Model

- RNN normal/obese group classifier result

DISCUSSION

Interpretation & Contribution

- **Establishing a Basis for Representing Food Schema in Vector Space**
  → Provides a reusable infrastructure for future food-schema research.
- **Automated Category-Tagging Pipeline**
  High prediction accuracy demonstrates that the embedding effectively captures association-based meaning structure.
  → Suggests the potential to replace manual tagging in large-scale free-association datasets.

Future Directions

- **Potential for Normal vs. Obese Group Classification**
  Validation performance fluctuated due to individual differences.
  → More participants would allow clearer group patterns and better classification performance.
- **Corpus Expansion**
  Expand the corpus with SNS posts and reviews
  → Enables more refined food-schema representations.