DISTRIBUTIONAL SEMANTICS FOR UNDERSTANDING SPOKEN MEAL DESCRIPTIONS

Mandy Korpusik, Calvin Huang, Michael Price, James Glass
MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA

Introduction

- Existing approaches for treating obesity are hampered by the lack of low-burden methods for tracking food intake.
- ***Goal***: *create a nutrition dialogue system that automatically extracts foods from a user's spoken meal log.*

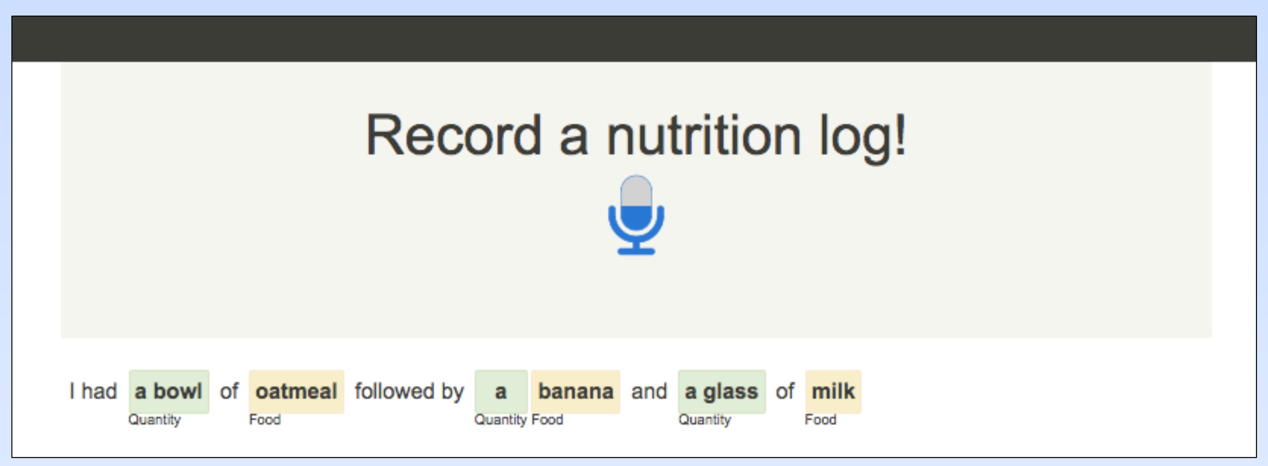


Fig. 1. The current system prototype.

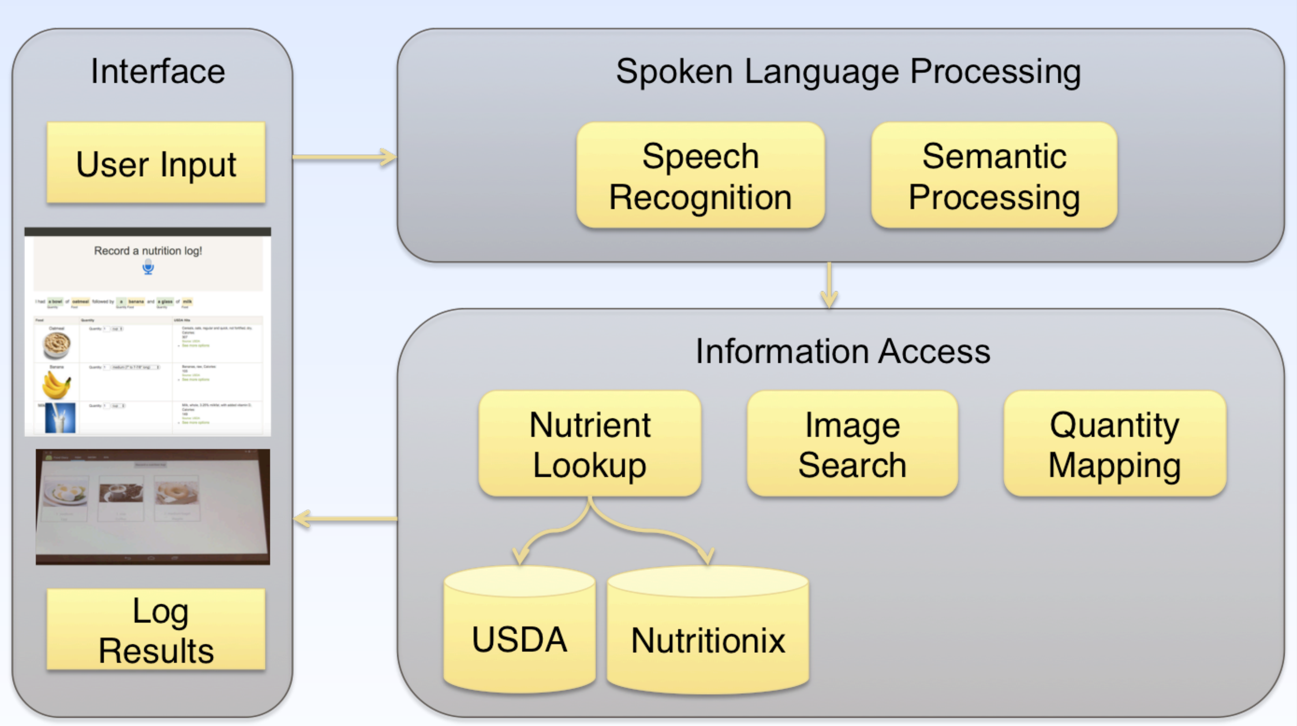


Fig. 2. The current system architecture.

Language Data

- We collected and labeled 10,000 breakfast/lunch/dinner/snack logs on Amazon Mechanical Turk (AMT).
- Three AMT tasks:
  - Writing meal descriptions
  - Labeling foods
  - Labeling properties (i.e., brand, quantity, and description)

Semantic Tagging

- ***Goal***: *label foods/properties in a meal log.*

I had **a bowl** (Quantity) of **Kellogg's** (Brand) **frosted** (Description) **flakes** (Food)

Classifiers

- Used conditional random field (CRF) model.
- Baseline features: n-grams, POS tags, food/ brand lexicon, and shape (e.g., capitalization).
- Distributional semantics features:
  - Dense word embeddings (word2vec)
  - Prototype similarity: cosine distance to 50 representative words for each label
  - Assigned word vectors to k-means clusters

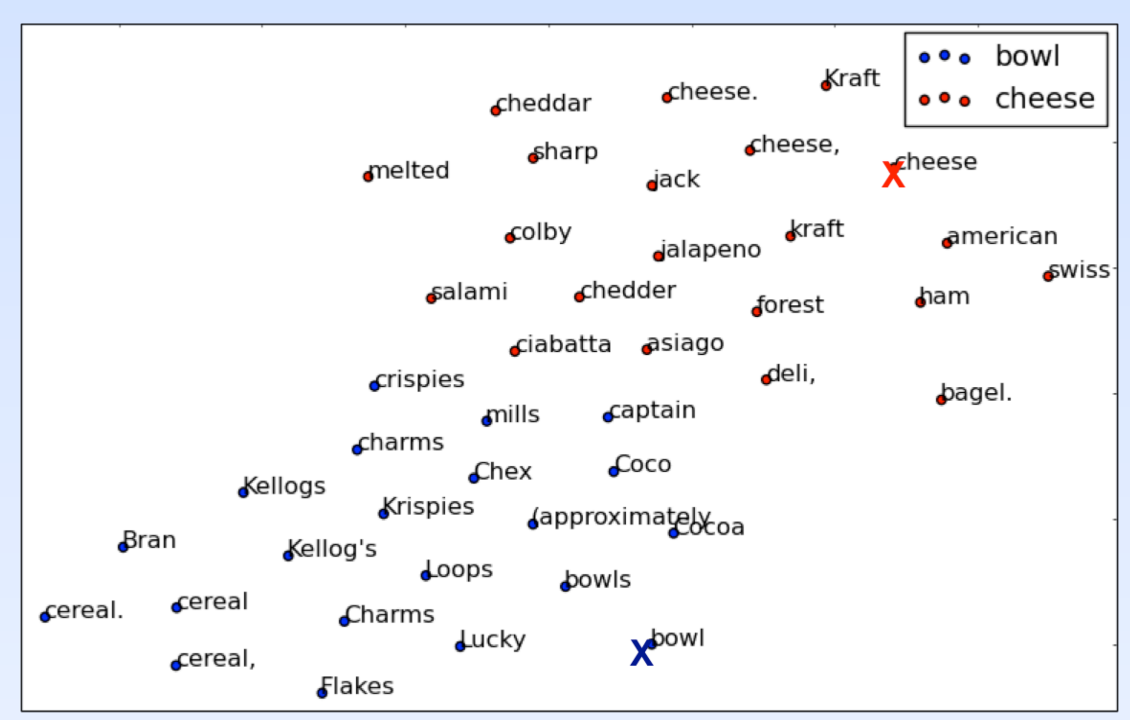


Fig. 3. 20 nearest words to "bowl"/"cheese" (vectors trained on nutrition data; plotted via t-SNE).

Model	Food	Brand	Num	Descr	Avg
Baseline	94.3	81.4	91.9	88.6	90.2
+ vectors	94.5	81.5	91.9	88.7	90.3
+ protos	94.9	82.4	**91.9**	89.0	90.7
+ shape	94.9	82.8	91.7	89.1	90.7
+ cluster	**95.0**	**82.8**	91.7	**89.1**	**90.8**

Table 1. F1 scores per label (except Other) with a CRF.

- **The combination of all baseline and distributional semantics features is best.**

Property Association

- ***Goal***: *associate properties with foods.*

I had **a bowl** (Quantity) of **cereal** (Food) and **two cups** (Quantity) of **milk** (Food)

Classifiers

- Trained classifiers to predict the most likely food for each property.
- Used features for each (food, property) pair:
  1. Property token
  2. Semantic tag of property token
  3. Distance between food and property
  4. Whether food is before/after property
- Experimented with random forest, naïve Bayes, and logistic regression. The random forest classifier performed best.
- Compared to previous BIO segmenting approach[1], and combined both methods.

[1] M. Korpusik, N. Schmidt, J. Drexler, S. Cyphers, and J. Glass. "Data Collection and Language Understanding of Food Descriptions" *Proc. SLT*, 2014.

Model	Prec	Recall	F1
Classifier (Oracle)	96.2	96.2	96.2
Segmenting (Oracle)	87.9	83.9	85.9
Combined (Oracle)	**96.5**	**96.5**	**96.5**
Classifier (Predict)	84.7	87.9	86.3
Segmenting (Predict)	**86.2**	81.0	83.5
Combined (Predict)	84.9	**88.2**	**86.5**

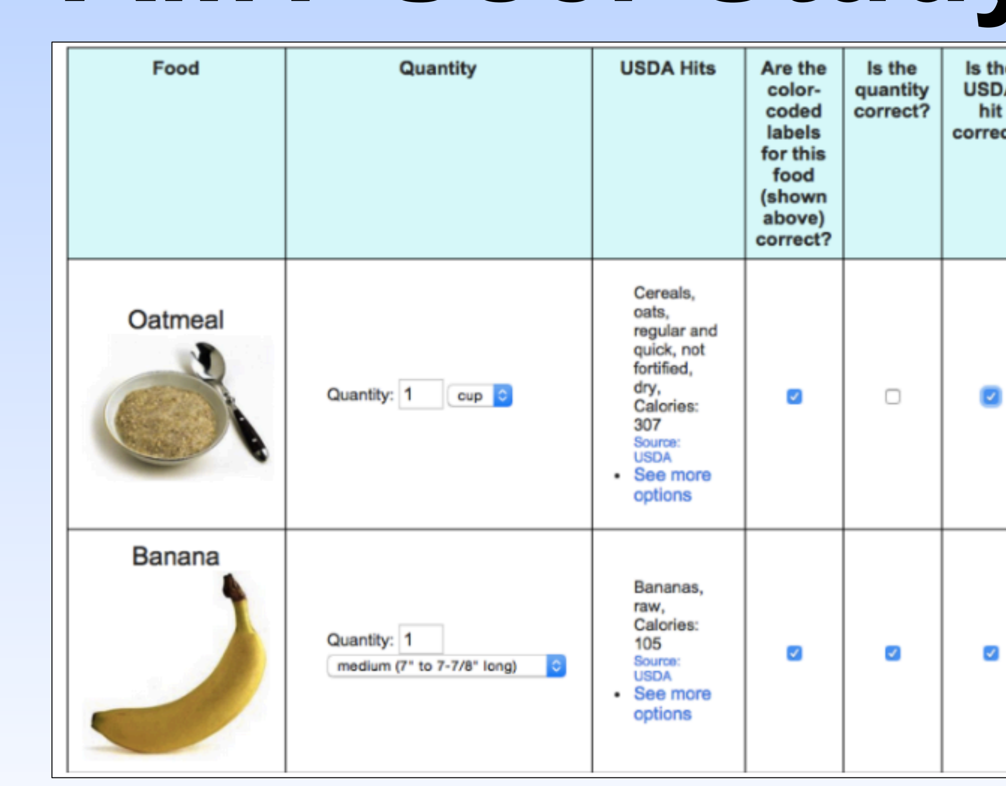
Table 2. Performance on property association task. Oracle experiments use AMT semantic tags (rather than CRF's predicted semantic tags).

- **The combination of segmenting and classifier approaches outperforms both methods individually.**

Speech Study

- Recorded 7,938 meal logs on AMT.
- Trained a speech recognizer in Kaldi.
- F1 scores on spoken test data:
  - Semantic tagging: 87.5
  - Property association: 86.0
- **Using spoken data did not greatly impact performance.**

AMT User Study



- Evaluated 437 meal descriptions.
- 83% semantic tagging accuracy.

Summary

- Significant improvement in semantic tagging with word vector features.
- Built a nutrition recognizer to evaluate performance on speech.
- **Ongoing work**: exploring neural methods and collecting more data.

Acknowledgments

- Rachael Naphtal and Patricia Saylor helped build the system.