DISTRIBUTIONAL SEMANTICS FOR UNDERSTANDING SPOKEN MEAL DESCRIPTIONS

Mandy Korpusik, Calvin Huang, Michael Price, James Glass

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA

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.

Introduction

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.

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

AMT User 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

Speech Study

- Using spoken data did not greatly impact performance.

Fig. 1. The current system prototype.

Semantic Tagging

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

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

Property Association

- Goal: associate properties with foods.

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.

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.

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Food</th>
<th>Brand</th>
<th>Num</th>
<th>Descr</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>94.3</td>
<td>81.4</td>
<td>91.9</td>
<td>88.6</td>
<td>90.2</td>
</tr>
<tr>
<td>+ vectors</td>
<td>94.5</td>
<td>81.5</td>
<td>91.9</td>
<td>88.7</td>
<td>90.3</td>
</tr>
<tr>
<td>+ protos</td>
<td>94.9</td>
<td>82.4</td>
<td>91.9</td>
<td>89.0</td>
<td>90.7</td>
</tr>
<tr>
<td>+ shape</td>
<td>94.9</td>
<td>82.8</td>
<td>91.7</td>
<td>89.1</td>
<td>90.7</td>
</tr>
<tr>
<td>+ cluster</td>
<td>95.0</td>
<td>82.8</td>
<td>91.7</td>
<td>89.1</td>
<td>90.8</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Prec</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier</td>
<td>96.2</td>
<td>96.2</td>
<td>96.2</td>
</tr>
<tr>
<td>Segmenting</td>
<td>87.9</td>
<td>83.9</td>
<td>85.9</td>
</tr>
<tr>
<td>Combined</td>
<td>96.5</td>
<td>96.5</td>
<td>96.5</td>
</tr>
<tr>
<td>Classifier</td>
<td>84.7</td>
<td>87.9</td>
<td>86.3</td>
</tr>
<tr>
<td>Segmenting</td>
<td>86.2</td>
<td>81.0</td>
<td>83.5</td>
</tr>
<tr>
<td>Combined</td>
<td>84.9</td>
<td>88.2</td>
<td>86.5</td>
</tr>
</tbody>
</table>

Fig. 2. The current system architecture.

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

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

Fig. 1. The current system prototype.

AMT User Study

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

Acknowledgments

- Rachael Naphtal and Patricia Saylor helped build the system.