Cognition-aware Cognate Detection

Diptesh Kanojia | Prashant K. Sharma | Sayali Ghodekar | Pushpak Bhattacharyya | Gholamreza Haffari | Malhar Kulkarni

European Chapter of the Association for Computational Linguistics
Cognate Detection: Motivation

- Cognates represent a large chunk of the shared vocabulary among language pairs.

- We conduct this experiment for an Indian language pair Hindi - Marathi, which is a known closely related pair.

- Previously, the task of Cognate Detection has shown to help the downstream tasks of Machine Translation via word alignment (Kondrak, 2005)

- Cognitive Psycholinguistic based features have also shown to improve various NLP tasks (Mishra et. al., 2016)
Cognition Aware Cognate Detection [1 /2]

Problem Statement

**Key Question:** Do cognitive (gaze) features help in cognate detection?

**GOALS**

- **Collect gaze behaviour data** for the task of identifying cognates vs. non-cognates for a sample set.
- **Extract gaze features** from the collected gaze data.
- **Predict gaze features** for the unseen samples.
- **Perform the task of cognate detection** over both sets.

**INPUT**

Cognate Challenge Dataset (Kanojia et. al., 2020) +
Traditional features +
Gaze data

**OUTPUT**

Cognates (1) / Non-Cognates (0)
Cognition Aware Cognate Detection [ 2 / 2]

- **Vector Representation:**
  - W1, W2, D1, D2, E1, E2
  - From Cognate Challenge Dataset
    - (Kanojia et. al., 2020)
- **Traditional features**
  - Phonetic, Lexical etc.
- **Gaze Features**
  - g1, g2, g3,.....gn
  - from collected data

**INPUT**

Vector Representation + Traditional features + Gaze data

**OUTPUT**

Cognates (1) / Non-Cognates (0)
Sub-Problem: Predicting Cognitive Features

**Problem Statement**

**GOAL**
- Using the collected gaze data, predict gaze features for the unseen samples of cognates and non-cognates.

- **Vector Representation:**
  - W1, W2, D1, D2, E1, E2

- **Traditional features**
  - Phonetic, Lexical etc.

- **Gaze Features**
  - g1, g2, g3,.....gn
  - from collected data

**INPUT**

Vector Representation
+ Traditional features
+ Gaze Features
(from collected data)

**OUTPUT**

Gaze Features
(on unseen data)

G1, G2, G3,.....Gn
Literature Survey

Cognate Detection

- **Using Phonetic Features**

- **Using Orthographic Features**

- **Using Cross-lingual Features**
  - Kanojia et. al. Utilizing cross-lingual word embeddings for Cognate Detection (COLING 2020)
Dataset Collection Setup

**Annotator Info**

- Nine annotators
- Native Marathi speakers (who understand Hindi)
- Education Level
  - At least College Graduates
- Experiments conducted with a host always at the side
- SR Research EyeLink 1000 (used at 500 Hz sampling rate)

To **verify the annotation quality** we observed two key aspects

- Annotation Precision (both individual and aggregate)
- Inter Annotator Agreement among our nine annotators (Fleiss Kappa Score)
Goal:
- Given cognate, and non-cognate pair along with their context (definition and example) collect gaze features for two hundred samples (100 +ve, 100 -ve).
Annotator Precision and Inter-annotator Agreement

<table>
<thead>
<tr>
<th>Annotator</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.99</td>
<td>0.975</td>
<td>0.965</td>
<td>0.995</td>
<td>0.995</td>
<td>0.99</td>
<td>0.975</td>
<td>0.99</td>
<td>0.98</td>
<td><strong>0.9839</strong></td>
</tr>
</tbody>
</table>

Cohen’s Kappa vs. Fleiss’ Kappa

Statistical literature observes that Cohen’s kappa is applicable to two annotators.

There are studies which use Cohen’s kappa for multiple annotators by computing a mean.

Fleiss’ kappa, however, allows multiple annotators, and categorical values to be taken into account.

We use Fleiss’ Kappa for statistical significance.
Cognate Detection with Gaze Features
**Proposed Model 1**: Neural Model for Cognition aware Cognate Detection

- **Feature Representation**: WordNet
- **Regression Model**: 
  - Input: `W1, D1, E1`
  - Output: Cognitive Features (Pred.)
- **Classification Model**: 
  - Input: Cognitive Features (Pred.)
  - Output: Cognates / Non-Cognates
  - `W1, W2`
- **Classification Model**: 
  - Input: Cognitive Features (Gaze)
  - Output: Cognates / Non-Cognates

**NOTE**:

- `W1, W2`: word pairs from two languages
- `D1, D2`: definition of `w1,w2`
- `E1, E2`: example of `w1, w2`

**Two Stages**:
- Stage-1: Cognitive Feature Prediction
- Stage-2: Classification model for Cognate Detection

- The regression and classification model are trained separately
## Results

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>Phonetic</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rama et. al., 2016 (D1+D2)</td>
<td>0.71 0.69 0.70</td>
<td>- - -</td>
</tr>
<tr>
<td>Kanojia et. al., 2019 (D1+D2)</td>
<td>- - -</td>
<td>0.76 0.72 0.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM</th>
<th>MUSE</th>
<th>VecMap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM (D1+D2)</td>
<td>0.83 0.71 0.77</td>
<td>0.72 0.68 0.70</td>
<td>0.70 0.65 0.67</td>
</tr>
<tr>
<td>LogisticRegression (D1+D2)</td>
<td>0.85 0.74 0.79</td>
<td>0.80 0.71 0.75</td>
<td>0.70 0.66 0.68</td>
</tr>
<tr>
<td>FFNN (D1 + D2)</td>
<td>0.82 0.84 <strong>0.83</strong></td>
<td>0.83 0.79 0.81</td>
<td>0.75 0.76 0.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM+Gaze</th>
<th>MUSE+Gaze</th>
<th>VecMap+Gaze</th>
<th>Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM (D2)</td>
<td>0.81 0.69 0.75</td>
<td>0.72 0.73 0.72</td>
<td>0.70 0.75 0.72</td>
<td>0.77 0.76 0.76</td>
</tr>
<tr>
<td>LogisticRegression (D2)</td>
<td>0.84 0.75 0.79</td>
<td>0.76 0.72 0.74</td>
<td>0.81 0.71 0.76</td>
<td>0.80 0.75 0.77</td>
</tr>
<tr>
<td>FFNN (D2)</td>
<td>0.83 0.85 <strong>0.84</strong></td>
<td>0.83 0.78 0.80</td>
<td>0.86 0.83 0.84</td>
<td>0.81 0.71 0.76</td>
</tr>
</tbody>
</table>

**Predicted Gaze Features On D1 (11652 samples) and Collected Gaze Features on D2 (200 samples)**

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM+Gaze</th>
<th>MUSE+Gaze</th>
<th>VecMap+Gaze</th>
<th>Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFNN (D1 + D2)</td>
<td>0.84 0.88 <strong>0.86</strong></td>
<td>0.85 0.78 0.81</td>
<td>0.83 0.85 0.84</td>
<td>0.77 0.76 0.76</td>
</tr>
<tr>
<td>FFNN (D1) [Only Predicted Gaze]</td>
<td>0.83 0.84 0.83</td>
<td>0.82 0.79 0.80</td>
<td>0.80 0.86 0.83</td>
<td>0.76 0.77 0.76</td>
</tr>
</tbody>
</table>
Model 1: Observation

- Our experiments shows that Introducing Gaze Features, results in improving cognate detection accuracy.
- Even on limited samples (1800 samples), our model shows improvement for the task of cognate detection
- Leveraging context information using neural architecture can help improving cognate detection accuracy.
Cognitive Features Prediction
**Proposed Model 2**: Neural Model for Cognitive Feature prediction

**Feature Representation**
- W1, D1, E1
- W2, D2, E2

**WordNet**

**Cognitive Features (Gaze)**

**Neural Regression Model**
- Linear layer
- Sigmoid
- Dropout (p=0.2)

**NOTE:**
- W1, W2: word pairs from two languages
- D1, D2: definition of w1, w2
- E1, E2: example of w1, w2

- **Single Stage**
- MSE loss

Presentation for EACL 2021 (19th-23rd April) || “Cognition-aware Cognate Detection”
Model 1: Results

1. AVERAGE FIXATION DURATION
2. AVERAGE SACCade AMPLITUDE
3. FIXATION COUNT
4. FIXATION DURATION MAX
5. FIXATION DURATION MIN
6. IA COUNT
7. RUN COUNT
8. SACCade COUNT
## Results

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>Phonetic</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rama et. al., 2016 (D1+D2)</td>
<td>0.71</td>
<td>0.69</td>
</tr>
<tr>
<td>Kanojia et. al., 2019 (D1+D2)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM</th>
<th>MUSE</th>
<th>VecMap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM (D1+D2)</td>
<td>0.83</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>LogisticRegression (D1+D2)</td>
<td>0.85</td>
<td>0.74</td>
<td>0.79</td>
</tr>
<tr>
<td>FFNN (D1 + D2)</td>
<td>0.82</td>
<td>0.84</td>
<td>0.83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM+Gaze</th>
<th>MUSE+Gaze</th>
<th>VecMap+Gaze</th>
<th>Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM (D2)</td>
<td>0.81</td>
<td>0.69</td>
<td>0.75</td>
<td>0.72</td>
</tr>
<tr>
<td>LogisticRegression (D2)</td>
<td>0.84</td>
<td>0.75</td>
<td>0.79</td>
<td>0.76</td>
</tr>
<tr>
<td>FFNN (D2)</td>
<td>0.83</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
</tr>
</tbody>
</table>

### Predicted Gaze Features On D1 (11652 samples) and Collected Gaze Features on D2 (200 samples)

<table>
<thead>
<tr>
<th>Feature Set →</th>
<th>XLM+Gaze</th>
<th>MUSE+Gaze</th>
<th>VecMap+Gaze</th>
<th>Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFNN (D1 + D2)</td>
<td>0.84</td>
<td>0.88</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>FFNN (D1) [Only Predicted Gaze]</td>
<td>0.83</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Future Investigation

**Cognate Detection**

**Using Phonetic Features**

**Using Orthographic Features**

**Using Cross-lingual Features**
- Kanojia et. al. Utilizing cross-lingual word embeddings for Cognate Detection (COLING 2020)

**Appending Cognitive Features**
- Kanojia et. al. Cognition-aware Cognate Detection (EACL 2021)
Future Work

- Cognate Detection using predicted gaze features on full corpus (Kanojia et. al. 2020 )
- Multi-Task Learning to predict gaze features and use it to predict a label for whether the words are cognates or not.
- Leveraging richer context representation for task of cognate detection.
- Predicting cognitive features for major NLP tasks: eg. Sentiment Analysis, Sarcasm Detection etc.
- Leveraging cognitive features for the task of word sense disambiguation.
References

Thank You!