A Novel Ensemble Method
for Named Entity Recognition
and Disambiguation
based on
Neural Network
“In the spring of 1800, Napoleon and his troops crossed the Swiss Alps”

Varying taxonomy of entity types for each extractor (PERSON, ORGANIZATION, LOCATION + others)
# Extractors

<table>
<thead>
<tr>
<th></th>
<th>Type recognition</th>
<th>Entity Disambiguation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alchemy API</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>Dandelion API</td>
<td>✗</td>
<td>✔ Wikipedia</td>
</tr>
<tr>
<td>DBSpotlight</td>
<td>✗</td>
<td>✔ DBpedia</td>
</tr>
<tr>
<td>TextRazor</td>
<td>✔</td>
<td>✔ Wikidata</td>
</tr>
<tr>
<td>Babelfy</td>
<td>✗</td>
<td>✔ DBpedia</td>
</tr>
<tr>
<td>MeaningCloud</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>ADEL</td>
<td>✔</td>
<td>✔ DBpedia</td>
</tr>
<tr>
<td>OpenCalais</td>
<td>✔</td>
<td>~</td>
</tr>
</tbody>
</table>
Ensemble strategies

Combine the result of different extractors and return the best answer.
10 extractors

Mappings to the NERD Ontology:

- Manual
  [Rizzo, 2011]
- Machine Learning (NERD-ML)
  [van Erp, 2013]

SVM or Naive Bayes perform the ensemble

5 extractors
(different ones with respect to NERD)

Mappings to a small typeset
{LOCATION, ORGANIZATION, PERSON}

MLP performs the ensemble
Our approach
Ensemble NERD

- No mapping required
  automatic alignment to the typeset of the ground truth (GT)

- Predict the most trustworthy extractor
  rather than the right type/entity

1. Surface form features
2. Type features
3. Entity features
4. Score features
4 KINDS OF FEATURE

\[ \mathbf{x} = \text{extractor} \]

\[ \mathbf{w} = \text{token (word)} \]

\[ \mathbf{e} = \text{named entity} \]
1. Surface Form Features

Text-related features

Specific corpus
(ex. OKE, AIDA, ...)

A Localized Wikipedia

\[ s^t \, = \, [\text{vector1}, \text{vector2}] \]
2. Type Features

Type taxonomy features

\[ t^e_x = \text{[ type vector]} \]

Any extractor-specific taxonomy
OR class hierarchy from Wikidata (NED only extractors)
2. Type Features

Type taxonomy features

\[ t_x^e = [ \text{type vector} ] \]

<table>
<thead>
<tr>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td><strong>Representation</strong></td>
</tr>
<tr>
<td>PERSON</td>
<td>001</td>
</tr>
<tr>
<td>ORGANIZATION</td>
<td>010</td>
</tr>
<tr>
<td>PLACE</td>
<td>100</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td><strong>Representation</strong></td>
</tr>
<tr>
<td>ACTOR</td>
<td>0001</td>
</tr>
<tr>
<td>MUSICIAN</td>
<td>0010</td>
</tr>
<tr>
<td>CITY</td>
<td>0100</td>
</tr>
<tr>
<td>MOUNTAIN</td>
<td>1000</td>
</tr>
</tbody>
</table>

**MUSICIAN** = 0010010

**PLACE** = 1000000

The dimensionality depends on the extractor
3. Entity Features

Similarity between 2 entities

\[ \text{Sim}(e_{x_1}, e_{x_2}) = \]

- URI MATCH
- LABELS SIMILARITY (Levenshtein)
- ABSTRACTS SIMILARITY (TF-IDF)
- OCCUPATION MATCH (Only for PERSON)
- STRUCTURAL SIMILARITY

Shortest distance (walk) between entities in Wikidata considering only subclass of (P279), instance of (P31), part of (P361)
3. Entity Features

Similarity between 2 entities

\[ f_e = \left\{ \text{Sim}(e_{x_1}, e_{x_2}), \text{Sim}(e_{x_1}, e_{x_3}), \text{Sim}(e_{x_1}, e_{x_4}), \ldots, \text{Sim}(e_{x_{n-1}}, e_{x_n}) \right\} \]

ex. TextRazor
ex. DBSpotlight
4. Score Features

\[ k^e_x = [\text{scores of the extractor } x] \]

The dimensionality depends on the extractor
2 NEURAL NETWORKS

Ensemble NN for Type Recognition
ENNTR

Ensemble NN for Disambiguation
ENND
1. ENNTR

Ensemble NN for Type Recognition

QUESTION
What is the right type?

NB OF RUNS
1 for each token
1. **ENNTR**

Ensemble NN for Type Recognition

- $H$ is the num of types in the GT
- $M_k$ is formed by $H$ neurons
- $M_k$ activation is linear
- **alignment block** strengths
  - it helps **avoiding local minima**
  - it doesn't privilege input with high dimensionality
  - it **aligns the type** between the extractor types and the GT types
## 1. ENNTR

Ensemble NN for Type Recognition

<table>
<thead>
<tr>
<th>token</th>
<th>alchemy output</th>
<th>alignment block output</th>
</tr>
</thead>
<tbody>
<tr>
<td>president</td>
<td>JobTitle</td>
<td>Role</td>
</tr>
<tr>
<td>eugenio</td>
<td>Person</td>
<td>Person</td>
</tr>
<tr>
<td>canfari</td>
<td>Person</td>
<td>Person</td>
</tr>
<tr>
<td>enrico</td>
<td>Person</td>
<td>Person</td>
</tr>
<tr>
<td>university</td>
<td>Organization</td>
<td>Organization</td>
</tr>
<tr>
<td>of</td>
<td>Organization</td>
<td>Organization</td>
</tr>
<tr>
<td>turin</td>
<td>Organization</td>
<td>Organization</td>
</tr>
<tr>
<td>foreign</td>
<td>JobTitle</td>
<td>Role</td>
</tr>
<tr>
<td>member</td>
<td>JobTitle</td>
<td>Role</td>
</tr>
</tbody>
</table>
1. ENNTR
Ensemble NN for Type Recognition

- The $O$ activations are linear
- At each neuron in $O$ corresponds a ground truth type
- To establish the type for a specific token, the type related with the highest neuron output value is chosen, if it greater than a threshold of 0.5
2. ENND

Ensemble NN for Disambiguation

Voting mechanism
2. ENND

Ensemble NN for Disambiguation

**QUESTION**

Is the candidate entity the right one?

1 = yes

0 = no

(sigmoid activation)

**NB OF RUNS**

1 for each candidate

(distinct extracted entities)
2. ENND

Ensemble NN for Disambiguation

1. Entity features
   similarity of the candidate to each extractor output

2. Type features
   same as ENTTR
2. ENND

Ensemble NN for Disambiguation

Diagram:
- **ENTITY**
  - Entity similarity features
  - Concatenate

- **TYPE**
  - Type features extractor 1
  - Type features extractor N
  - Dense
  - Dense

- 2 dense layers
  - **selu** activation

Output layer (O)
2. ENND

Ensemble NN for Disambiguation

**Voting mechanism**
for each token $t$, after $N$ network run

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Output neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>$o_1$</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>$c_N$</td>
<td>$o_N$</td>
</tr>
</tbody>
</table>

If $O_{\text{max}} > 0.5$, valid candidate; otherwise, rejected candidate.
RESULTS
Ground Truth Corpora

● **OKE2016** 4 types
  ![DBpedia](https://dbpedia.org)

● **AIDA/CoNLL** used only for NED
  ![DBpedia](https://dbpedia.org)

● **NexGenTV corpus** (1) 13 types
  ![WIKIDATA](https://w.warin.com)

(1) [http://enerd.eurecom.fr/data/training_data/nexgen_tv_corpus/](http://enerd.eurecom.fr/data/training_data/nexgen_tv_corpus/)

Metrics

● **MCU-7 test scoring**
  Type recognition

● **Gerbil scores**
  Entity disambiguation
Type Recognition Evaluation

<table>
<thead>
<tr>
<th>EXTRACTOR</th>
<th>p OKE 2016</th>
<th>r OKE 2016</th>
<th>f1 OKE 2016</th>
<th>p NexGenTV</th>
<th>r NexGenTV</th>
<th>f1 NexGenTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>adel</td>
<td>0.85</td>
<td>0.83</td>
<td>0.84</td>
<td>0.83</td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>alchemy</td>
<td>0.92</td>
<td>0.86</td>
<td>0.88</td>
<td>0.97</td>
<td>0.81</td>
<td>0.87</td>
</tr>
<tr>
<td>babelfy</td>
<td>0.79</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
<td>0.59</td>
<td>0.65</td>
</tr>
<tr>
<td>dandelion</td>
<td>0.83</td>
<td>0.75</td>
<td>0.78</td>
<td>0.69</td>
<td>0.42</td>
<td>0.51</td>
</tr>
<tr>
<td>dbspotlight</td>
<td>0.77</td>
<td>0.52</td>
<td>0.60</td>
<td>0.61</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>meaning cloud</td>
<td>0.78</td>
<td>0.69</td>
<td>0.72</td>
<td>0.87</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>opencalais</td>
<td>0.71</td>
<td>0.68</td>
<td>0.69</td>
<td>0.90</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>textrazor</td>
<td>0.81</td>
<td>0.74</td>
<td>0.77</td>
<td>0.80</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>ensemble</td>
<td>0.95</td>
<td>0.92</td>
<td>0.94</td>
<td>0.98</td>
<td>0.87</td>
<td>0.92</td>
</tr>
<tr>
<td>ensemble (TYPE only)</td>
<td>0.92</td>
<td>0.84</td>
<td>0.88</td>
<td>0.93</td>
<td>0.85</td>
<td>0.89</td>
</tr>
<tr>
<td>ensemble (SURFACE FORM only)</td>
<td>0.52</td>
<td>0.48</td>
<td>0.50</td>
<td>0.56</td>
<td>0.50</td>
<td>0.53</td>
</tr>
<tr>
<td>ensemble (ENTITY only)</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.47</td>
<td>0.45</td>
<td>0.46</td>
</tr>
<tr>
<td>ensemble (SCORE only)</td>
<td>0.40</td>
<td>0.36</td>
<td>0.38</td>
<td>0.40</td>
<td>0.38</td>
<td>0.39</td>
</tr>
</tbody>
</table>
# NED Evaluation

<table>
<thead>
<tr>
<th>EXTRACTOR</th>
<th>OKE 2016</th>
<th></th>
<th>AIDA</th>
<th></th>
<th>NexGenTV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p</td>
<td>r</td>
<td>f1</td>
<td>p</td>
<td>r</td>
<td>f1</td>
</tr>
<tr>
<td>babelfy</td>
<td>0.64</td>
<td>0.47</td>
<td>0.54</td>
<td>0.70</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>dandelion</td>
<td>0.77</td>
<td>0.48</td>
<td>0.59</td>
<td>0.66</td>
<td>0.34</td>
<td>0.45</td>
</tr>
<tr>
<td>dbspotlight</td>
<td>0.53</td>
<td>0.30</td>
<td>0.39</td>
<td>0.65</td>
<td>0.36</td>
<td>0.47</td>
</tr>
<tr>
<td>textrazor</td>
<td>0.78</td>
<td>0.40</td>
<td>0.53</td>
<td>0.57</td>
<td>0.53</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>ensemble</strong></td>
<td>0.88</td>
<td>0.52</td>
<td>0.66</td>
<td>0.79</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>ensemble (ENTITY only)</strong></td>
<td>0.80</td>
<td>0.47</td>
<td>0.59</td>
<td>0.60</td>
<td>0.50</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>ensemble (TYPE only)</strong></td>
<td>0.45</td>
<td>0.38</td>
<td>0.41</td>
<td>0.52</td>
<td>0.45</td>
<td>0.48</td>
</tr>
</tbody>
</table>
NED Evaluation

<table>
<thead>
<tr>
<th>EXTRACTOR</th>
<th>OKE 2016</th>
<th></th>
<th>AIDA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p</td>
<td>r</td>
<td>f1</td>
<td>p</td>
</tr>
<tr>
<td>agdistis</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.58</td>
</tr>
<tr>
<td>aida</td>
<td>0.63</td>
<td>0.41</td>
<td>0.49</td>
<td>0.00</td>
</tr>
<tr>
<td>dexter</td>
<td>0.92</td>
<td>0.29</td>
<td>0.44</td>
<td>0.38</td>
</tr>
<tr>
<td>fox</td>
<td>0.77</td>
<td>0.35</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td>freme ner</td>
<td>0.57</td>
<td>0.21</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>kea</td>
<td>0.67</td>
<td>0.61</td>
<td>0.64</td>
<td>0.56</td>
</tr>
<tr>
<td>pboh</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>ensemble</strong></td>
<td><strong>0.88</strong></td>
<td><strong>0.52</strong></td>
<td><strong>0.66</strong></td>
<td><strong>0.60</strong></td>
</tr>
</tbody>
</table>

extractors in GERBIL
Conclusions - Main Findings

**NER: ENNTR**
- The NN finds the most trustworthy extractor for each case
- The ensemble method **outperforms** the single extractors (F1)
- The most useful features are the **type** ones
- Removing the type features, the ensemble method reaches different minima and outperforms the single extractors

**NED: ENND**
- The NN predicts if a result is good or not + voting system
- The ensemble method **outperforms** the single extractors (F1)
- The most useful features are the **entity** ones
- The **type** improves of **10%** the final F1 score
Future Work

- add new extractors to the ensemble (e.g. Spacy, ELMo, Flair)
- use LSTM/BiLSTM for surface form features
- use Wikidata embedding
- enhance the NexGenTV corpus training data
Thank you!

code: /D2KLab/ensemble-nerd

this presentation: https://goo.gl/5eHhUc