Ellen Riloff: Hey Hal, what happens if there are classes in the new domain that don't exist in the old domain?

Hal (c. 2007): Meh? No clue. Does that really happen in practice?

Language does have many flavors!

- Can you guess what domain each of these sentences is drawn from?

<table>
<thead>
<tr>
<th>Domain</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>Many factors contributed to the French and Dutch objections to the proposed EU constitution</td>
</tr>
<tr>
<td>Parliament</td>
<td>Please rise, then, for this minute's silence</td>
</tr>
<tr>
<td>Medical</td>
<td>Latent diabetes mellitus may become manifest during thiazide therapy</td>
</tr>
<tr>
<td>Science</td>
<td>Statistical machine translation is based on sets of text to build a translation model</td>
</tr>
<tr>
<td>Step-mother</td>
<td>I forgot to mention in yesterday's post that I also trimmed an overgrown HUGE hedge that spams the entire length of the front of my house and is about 3' accrossed.</td>
</tr>
</tbody>
</table>
S4 taxonomy of adaptation effects

- **Seen**: Never seen this word before
  - News to medical: “diabetes mellitus”

- **Sense**: Never seen this word used in this way
  - News to technical: “monitor”

- **Score**: The wrong output is scored higher
  - News to medical: “manifest”

- **Search**: Decoding/search erred
S4 applied to “easy” NLP problems...

Part of Speech Tagging

Inside = Old domain

Outside = New domain

Seen  Sense  Score
S4 applied to “easy” NLP problems...

Part of Speech Tagging

Shallow Parsing

Named Entity Recognition

Inside = Old domain

Outside = New domain
Is this a problem for harder tasks?

Translate

[Google Translate interface with Translate button and translation examples]

Kitchenaid in perfect condition
Anti Obama still perfect
Translating across domains is hard

<table>
<thead>
<tr>
<th>Old Domain (Parliament)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
</tr>
<tr>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td><strong>System</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
</tr>
<tr>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td><strong>System</strong></td>
</tr>
</tbody>
</table>

**Key Question:** What went wrong?
Senses are domain/language specific
Domain Shift Setting

**Old domain:** Hansard parliamentary proceedings

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Tokens</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>fr</td>
<td>162m</td>
<td>192k</td>
</tr>
<tr>
<td>en</td>
<td>145m</td>
<td>187k</td>
</tr>
</tbody>
</table>

French to English phrase-based translation system

Compare old domain model

With mixed old+new model as pseudo-oracle
## New Domain Datasets

<table>
<thead>
<tr>
<th>Domain</th>
<th>Sentences</th>
<th>Tokens</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>135k</td>
<td>fr 4m</td>
<td>63k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>en 3m</td>
<td>52k</td>
</tr>
<tr>
<td>Medical</td>
<td>472k</td>
<td>fr 7m</td>
<td>35k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>en 6m</td>
<td>30k</td>
</tr>
<tr>
<td>Science</td>
<td>139k</td>
<td>fr 4m</td>
<td>118k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>en 4m</td>
<td>114k</td>
</tr>
<tr>
<td>Subtitles</td>
<td>19,000k</td>
<td>fr 155m</td>
<td>362k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>en 174m</td>
<td>293k</td>
</tr>
<tr>
<td>Old domain</td>
<td>8,000k</td>
<td>fr 162m</td>
<td>192k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>en 145m</td>
<td>187k</td>
</tr>
</tbody>
</table>
SenseSpotting

Never let your parallel data tie you to an old domain

- Spotting new **words** is easy
- Spotting words with new **translations** is hard

- Binary classification problem:
  - +ve: French token has previously unseen sense
  - -ve: French token is used in a known way

- 16 fold cross validation *at the type level*
  - *Never* test on a word type seen in training!
  - train on *mode*, *administration*; test on *rapport*
Automatically identifying new senses

- Context + existence of translations in comparable data

is a window of opportunity to have a window of opportunity in the run up to, we run the risk via une fenêtre insérée. vers ma fenêtre ou vers voulons pas courir le risque, sans courir le risque

dans la fenêtre. cet

courir not found

e ne pouvez.exécuter que les pour l' exécuter elle va
New Sense Indicators

New senses alter corpus-level word frequency

New senses alter document-level context
  topic distribution

New senses alter local context
  n-gram language model
  distributional similarity
  context-dependent translation model
SenseSpotting Results

Area Under the ROC Curve

- **Medical**
  - All Features: Blue
  - Only Token: Red
  - Only Type: Yellow
  - Random: Maroon
  - Constant: Green

- **Subtitles**
  - All Features: Blue
  - Only Token: Red
  - Only Type: Yellow
  - Random: Maroon
  - Constant: Green

- **Science**
  - All Features: Blue
  - Only Token: Red
  - Only Type: Yellow
  - Random: Maroon
  - Constant: Green

Note: The area under the ROC curve is given for each category, with values of 45%, 23k tokens, 24%, 8k tokens, and 52%, 35k tokens.
Indicators to reach peak performance

New senses alter corpus-level word frequency
New senses alter document-level context
  topic distribution
New senses alter local context
  n-gram language model
  distributional similarity
  context-dependent translation model

Computed at both type and token levels
SenseSpotting summary

new task motivated by cross-domain machine translation errors
free token-level annotation from parallel text
minimal new domain parallel text required
AUC as high as 80%
on word types never seen during training
requires both type and token level indicators
Simultaneously solving seen+sense

• Idea:
  • We have good knowledge of translations in the old domain
  • We have good knowledge of raw word frequencies in a new domain in each language individually
  • Can we “nudge” the translation probabilities to match these raw frequencies

• Assumptions:
  • Old domain parallel data
  • New domain comparable data
Marginal matching for “sense” errors

Given:
- A joint $p(x,y)$ in the old domain
- Marginals $q(x)$ and $q(y)$ in the new domain

Recover:
- Joint $q(x,y)$ in the new domain

We formulate as a L1-regularized linear program

Easier: many $q(x)$ and $q(y)$s
Marginal matching

<table>
<thead>
<tr>
<th></th>
<th>house</th>
<th>place</th>
<th>pregnant</th>
<th>dress</th>
</tr>
</thead>
<tbody>
<tr>
<td>enceinte</td>
<td>0.30</td>
<td>0.40</td>
<td>0.10</td>
<td>0</td>
</tr>
<tr>
<td>habiller</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ q^{\text{old}}(s) \]

<table>
<thead>
<tr>
<th></th>
<th>house</th>
<th>place</th>
<th>pregnant</th>
<th>dress</th>
<th>girl</th>
</tr>
</thead>
<tbody>
<tr>
<td>enceinte</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>habiller</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ q(s) \]

- (a) OLD-Domain Joint
- (b) NEW-Domain Marginals
- (c) Inferred NEW-Domain Joint

Matched Marginals
Additional features

Sparsity: # of non-zero entries should be small

Distributional: document co-occurrence $\leftrightarrow$ translation pair

Spelling: Low edit dist $\leftrightarrow$ translation pair

Frequency: Rare words align to rare words; common words align to common words
Marginal matching details

\[ \Omega(p) \] : regularization term

\[ f(p) \] : edit distance penalty

\[ p^{\text{new}} = \arg \min_p \| p - p^{\text{old}} \|_1 + \Omega(p) + f(p) \]

subject to:

\[ \sum_{s,t} p(s,t) = 1, \quad p(s,t) \geq 0 \]

\[ \sum_s p(s,t) = q(t), \quad \sum_t p(s,t) = q(s) \]
## Example Learned Translations

<table>
<thead>
<tr>
<th>French</th>
<th>Correct</th>
<th>Learned Translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>cisaillement</td>
<td>shear</td>
<td>viscous crack shear</td>
</tr>
<tr>
<td>chromosomes</td>
<td>chromosomes</td>
<td>chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>chromosome</td>
</tr>
<tr>
<td></td>
<td></td>
<td>chromosomal</td>
</tr>
<tr>
<td>caractérisation</td>
<td>characterization</td>
<td>characterization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>characteristic $\pi$</td>
</tr>
<tr>
<td>araignées</td>
<td>spiders</td>
<td>spiders ant spider</td>
</tr>
<tr>
<td>tiges</td>
<td>stems</td>
<td>usda centimeters flowering</td>
</tr>
</tbody>
</table>
BLEU Scores

Baseline
Discussion

- Machine translation adaptation has lots of challenges:
  - New words appear all the time
  - Words gain new senses all the time
  - ... and the usual (new scores)
- S4 as a (qualitative) measure of divergence
- SenseSpotting helps when new classes arise
- Marginal matching: learning on comparable data

Thanks to:
- Fabienne Braune
- Marine Carpuat
- Ann Clifton
- Alex Fraser
- Katharine Henry
- Ann Irvine
- Jagadeesh Jagarlamudi
- John Morgan
- Chris Quirk
- Majid Razmara
- Rachel Rudinger
- Aleš Tamchyna

Thanks!  Questions?