Examining Temporality in Document Classification

Xiaolei Huang  Michael J. Paul
University of Colorado Boulder
Examining Temporality in Document Classification
or
Why is my classifier getting worse over time?
Why is my classifier getting worse?

• The data distribution has changed…
  • Is there anything systematic about how it changes?
  • Is there anything we can do to adapt to temporal changes?
Experiments

Two types of time periods:

• Seasonal
  • Repeat across years
    (e.g., time of year)

• Non-seasonal
  • No repetition
    (e.g., spans of years)
Experiments

- Binary classification
  - Logistic regression, n-gram features
- Six datasets, each grouped into 4-6 time periods

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time intervals (non-seasonal)</th>
<th>Time intervals (seasonal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews (hotels)</td>
<td>2005-08, 2009-11, 2012-14, 2015-17</td>
<td>Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec</td>
</tr>
<tr>
<td>Reviews (restaurants)</td>
<td>2005-08, 2009-11, 2012-14, 2015-17</td>
<td>Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec</td>
</tr>
<tr>
<td>Twitter (vaccines)</td>
<td>2013, 2014, 2015, 2016</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec</td>
</tr>
</tbody>
</table>
Why is my classifier getting worse?

• The data distribution has changed…
  • Is there anything systematic about how it changes?
  • Is there anything we can do to adapt to temporal changes?
RQ1: How does performance vary?

Analysis:

• Train and test on each time period
  • Measure how performance drops when the test period is different
• Balanced so each time period has same # of documents
RQ1: How does performance vary?
RQ1: How does performance vary?
RQ1: How does performance vary?

Yelp reviews are getting more informative over time?
RQ1: How does performance vary?

Takeaways:

• This type of analysis can reveal characteristics of corpus
• Unanswered: why does performance vary?
Why is my classifier getting worse?

• The data distribution has changed…
  • Is there anything systematic about how it changes?
  • Is there anything we can do to adapt to temporal changes?
RQ2: Can we adapt to temporal variations?

Idea:

• Address this as a **domain adaptation** problem
• Treat explicitly-defined time periods as domains
RQ2: Can we adapt to temporal variations?

Approach:

• Feature augmentation method from Daumé III (2007)
RQ2: Can we adapt to temporal variations?

Approach:

• Feature augmentation method from Daumé III (2007)
RQ2: Can we adapt to temporal variations?

Domain-specific copies of the feature set:

<table>
<thead>
<tr>
<th>General</th>
<th>Jan-Mar</th>
<th>Apr-Jun</th>
<th>Jul-Sep</th>
<th>Oct-Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="General" /></td>
<td><img src="image2" alt="Jan-Mar" /></td>
<td><img src="image3" alt="Apr-Jun" /></td>
<td><img src="image4" alt="Jul-Sep" /></td>
<td><img src="image5" alt="Oct-Dec" /></td>
</tr>
</tbody>
</table>
RQ2: Can we adapt to temporal variations?
RQ2: Can we adapt to temporal variations?

• Straightforward to apply to seasonal features:

<table>
<thead>
<tr>
<th>Data (Seasonal)</th>
<th>Baseline</th>
<th>Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews (music)</td>
<td>.901</td>
<td>.919</td>
</tr>
<tr>
<td>Reviews (hotels)</td>
<td>.867</td>
<td>.881</td>
</tr>
<tr>
<td>Reviews (restaurants)</td>
<td>.874</td>
<td>.898</td>
</tr>
<tr>
<td>News (economy)</td>
<td>.782</td>
<td>.782</td>
</tr>
<tr>
<td>Twitter (vaccines)</td>
<td>.881</td>
<td>.880</td>
</tr>
</tbody>
</table>
RQ2: Can we adapt to temporal variations?

- How to use in non-seasonal settings?
RQ2: Can we adapt to temporal variations?

• How to use in non-seasonal settings?
  • Separately weigh domain-specific features

General  2012  2013  2014  2015
RQ2: Can we adapt to temporal variations?

• How to use in non-seasonal settings?
  • During training: weigh domain-specific features differently
  • Can also combine with seasonal domains
    • 3 copies of each feature (general, year-specific, season-specific)
  • Simulating performance on future data:
    • Train in initial time periods
    • Tune on second-to-last period
    • Test on final time period
RQ2: Can we adapt to temporal variations?

- How to use in non-seasonal settings?

<table>
<thead>
<tr>
<th>Data (Non-seasonal)</th>
<th>Baseline</th>
<th>Adaptation</th>
<th>Adapt.+seasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews (music)</td>
<td>.895</td>
<td>.924</td>
<td>.910</td>
</tr>
<tr>
<td>Reviews (hotels)</td>
<td>.886</td>
<td>.892</td>
<td>.920</td>
</tr>
<tr>
<td>Reviews (restaurants)</td>
<td>.831</td>
<td>.879</td>
<td>.889</td>
</tr>
<tr>
<td>News (economy)</td>
<td>.763</td>
<td>.780</td>
<td>.859</td>
</tr>
<tr>
<td>Politics (platforms)</td>
<td>.661</td>
<td>.665</td>
<td>n/a</td>
</tr>
<tr>
<td>Twitter (vaccines)</td>
<td>.910</td>
<td>.903</td>
<td>.920</td>
</tr>
</tbody>
</table>
RQ2: Can we adapt to temporal variations?

Takeaways:

• Simple-to-implement adaptation can make classifiers more robust across time

• Suggestion: tune hyperparameters on heldout data from the chronological end of your corpus (cf. cross-validation)
  • Can lead to better performance on future data
Thank you!

Questions?

• Code: https://github.com/xiaoleihuang/Domain_Adaptation_ACL2018