

Examining Temporality in Document Classification

Xiaolei Huang

Michael J. Paul

University of Colorado Boulder



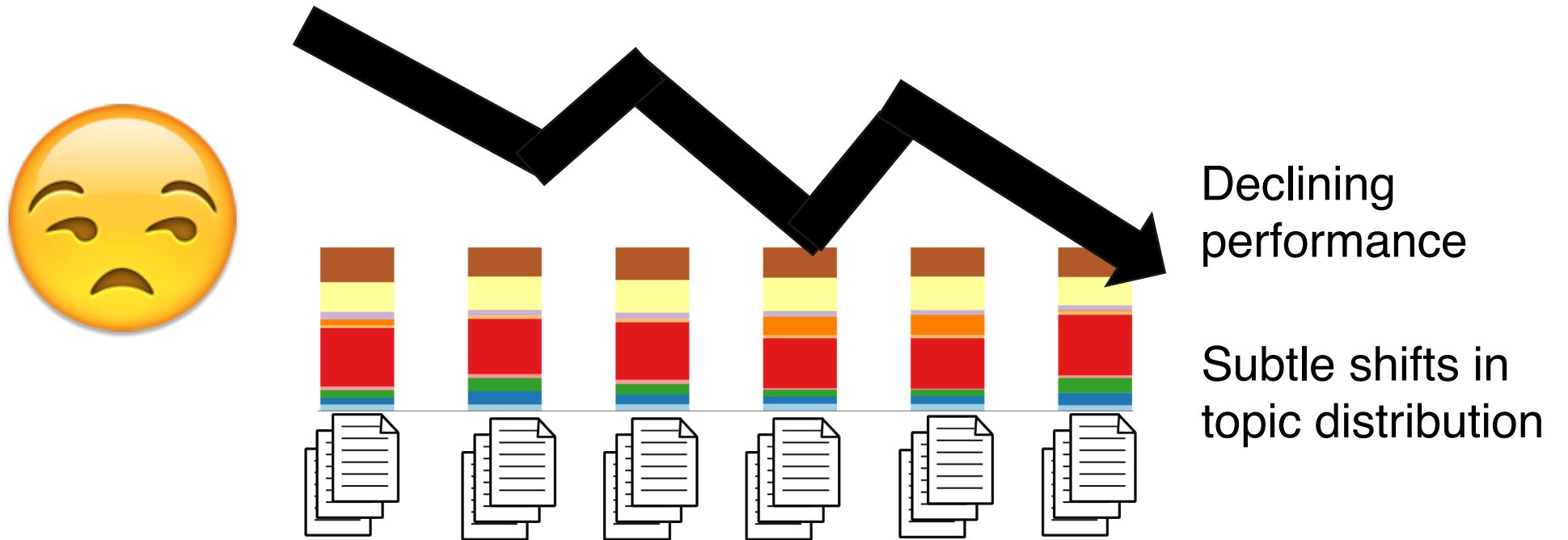
Examining Temporality in Document Classification

or

Why is my classifier getting worse over time?

Why is my classifier getting worse?

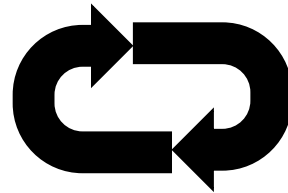
- The data distribution has changed...
 - Is there anything systematic about how it changes?
 - Is there anything we can do to adapt to temporal changes?



Experiments

Two types of time periods:

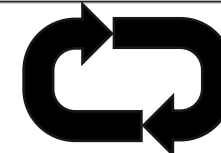
- Seasonal
 - Repeat across years
(e.g., time of year)
- Non-seasonal
 - No repetition
(e.g., spans of years)



Experiments

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

Dataset	Time intervals (non-seasonal)	Time intervals (seasonal)
Reviews (music)	1997-99, 2000-02, 2003-05, 2006-08, 2009-11, 2012-14	Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec
Reviews (hotels)	2005-08, 2009-11, 2012-14, 2015-17	Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec
Reviews (restaurants)	2005-08, 2009-11, 2012-14, 2015-17	Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec
News (economy)	1950-70, 1971-85, 1986-2000, 2001-14	Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec
Politics (platforms)	1948-56, 1960-68, 1972-80, 1984-92, 1996-2004, 2008-16	n/a
Twitter (vaccines)	2013, 2014, 2015, 2016	Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec



Why is my classifier getting worse?

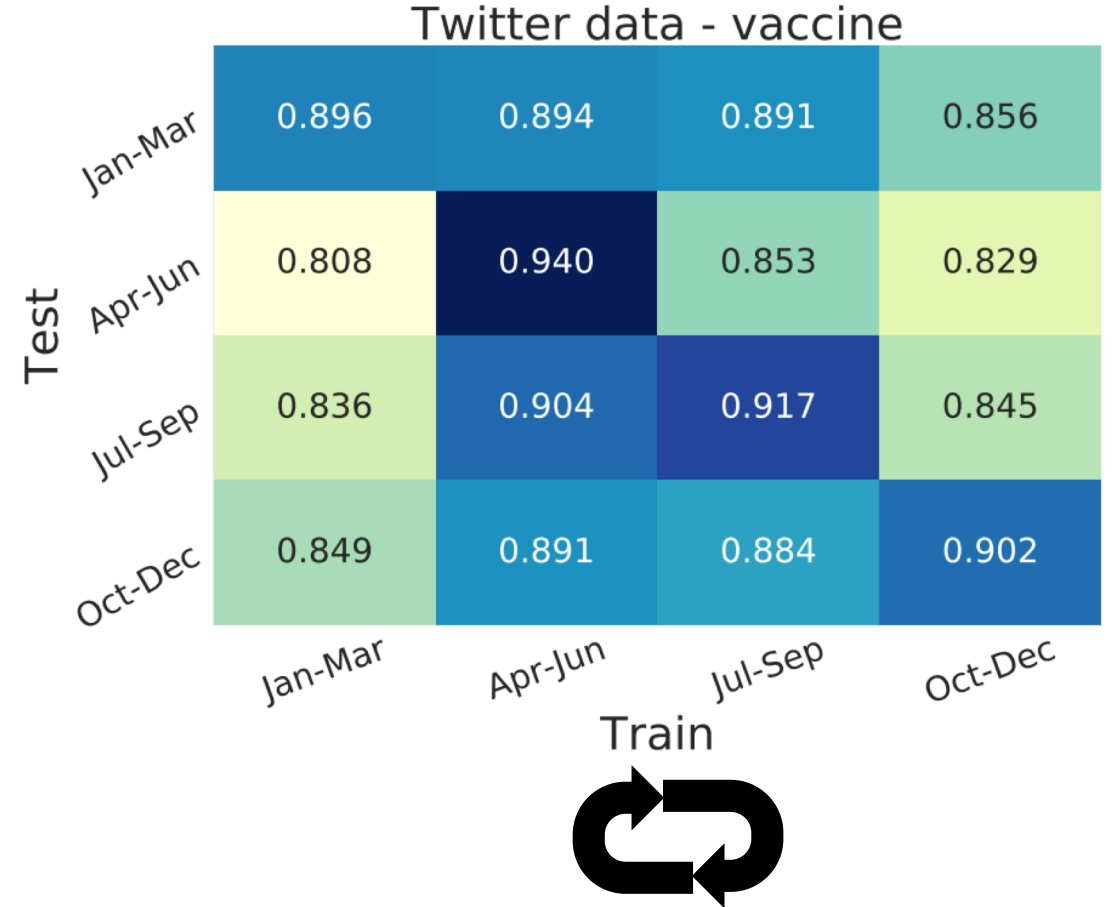
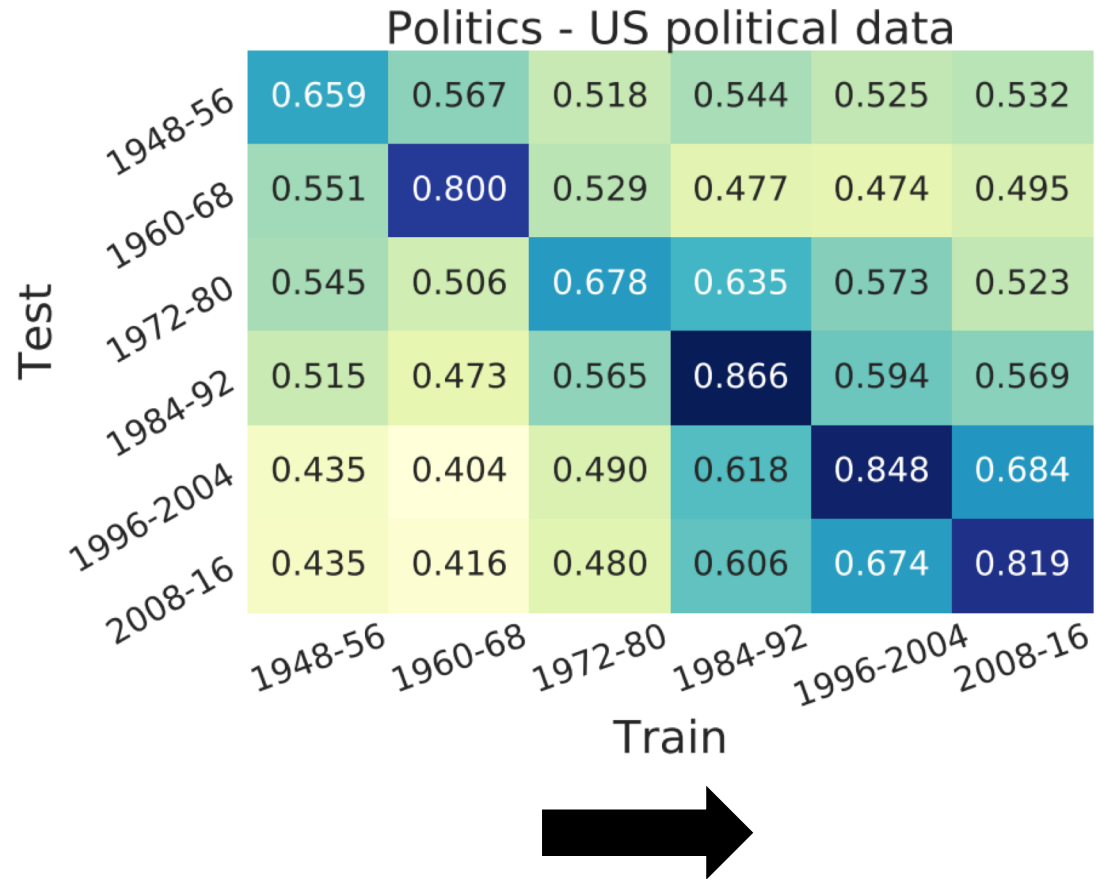
- The data distribution has changed...
 - **Is there anything systematic about how it changes?**
 - Is there anything we can do to adapt to temporal changes?

RQ1: How does performance vary?

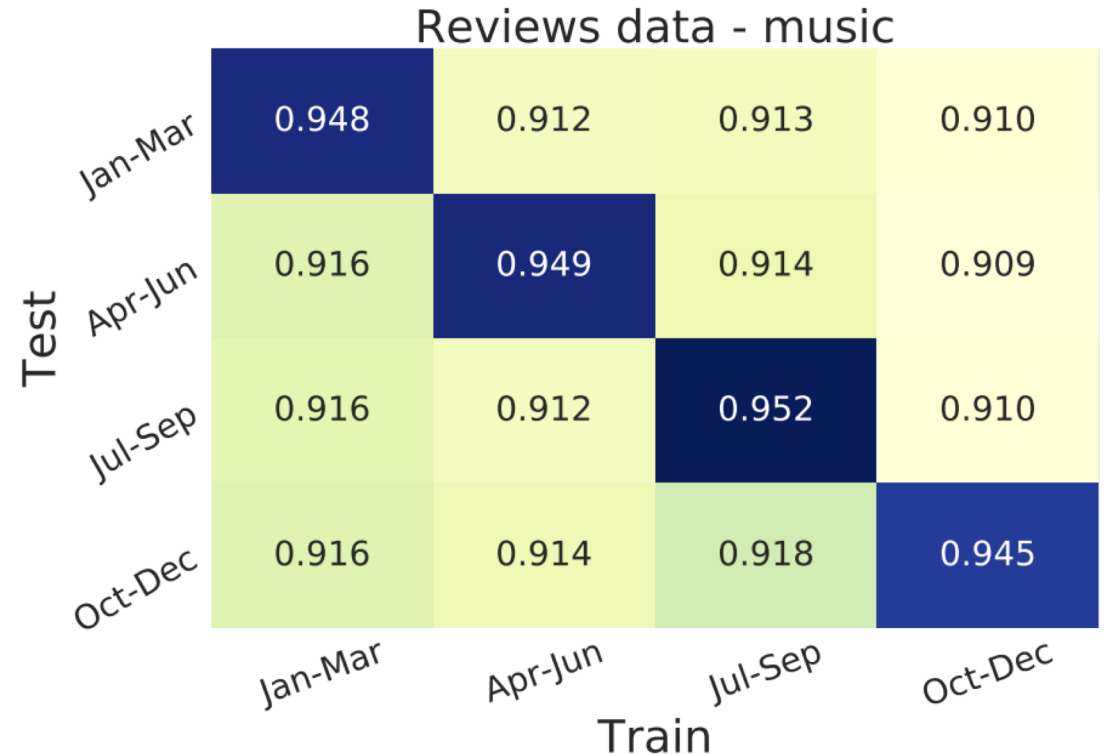
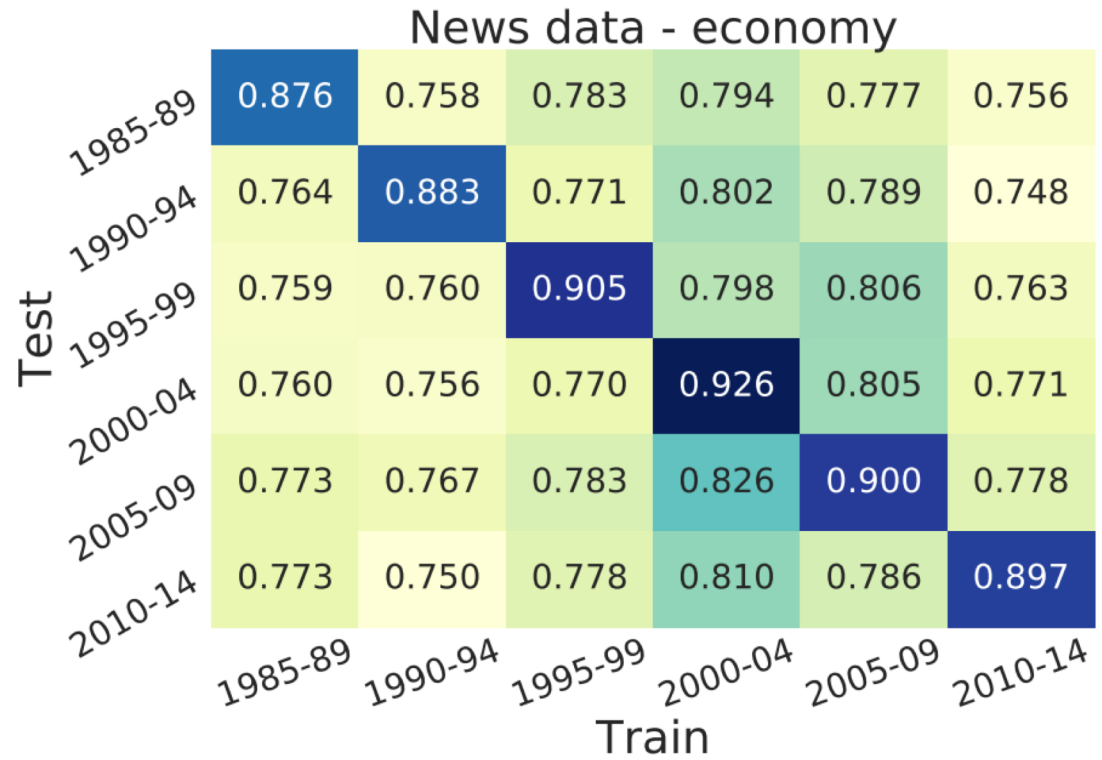
Analysis:

- 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

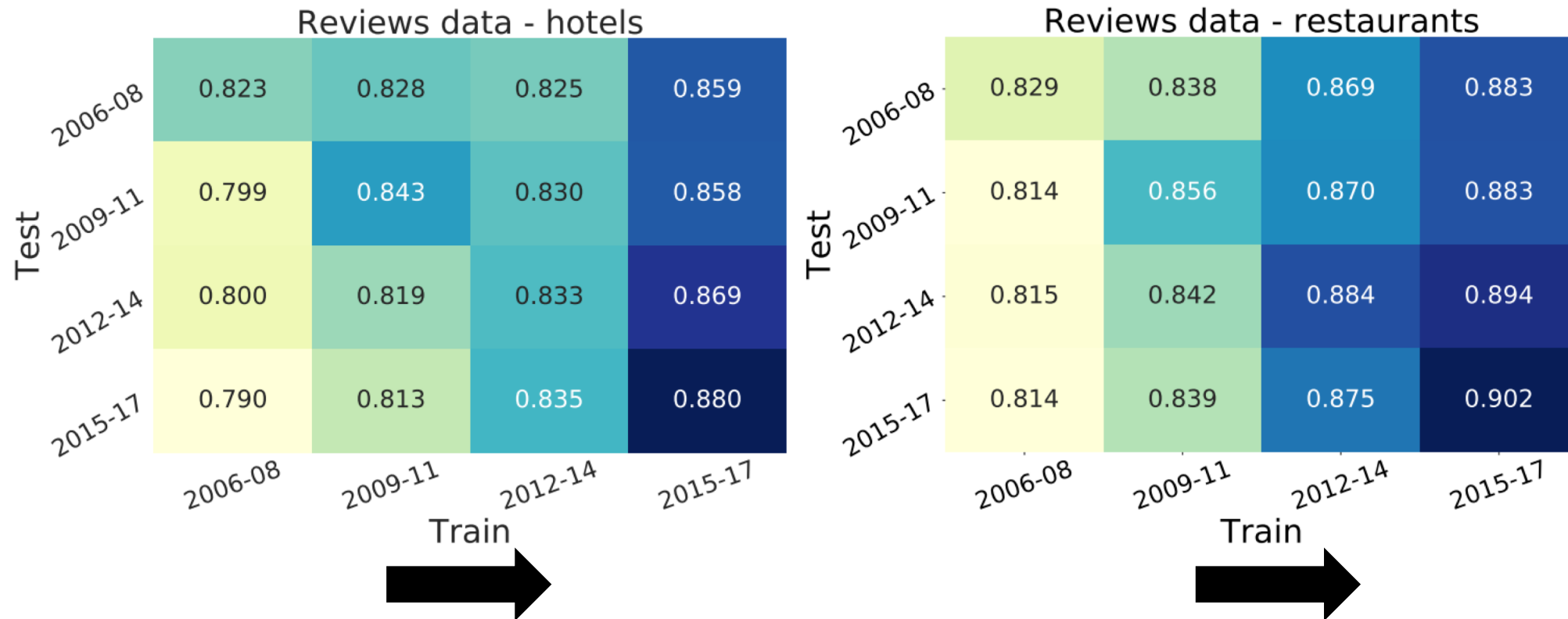
RQ1: How does performance vary?



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)



Photo via @ChrisVVarren

RQ2: Can we adapt to temporal variations?

Domain-specific copies of the feature set:

General



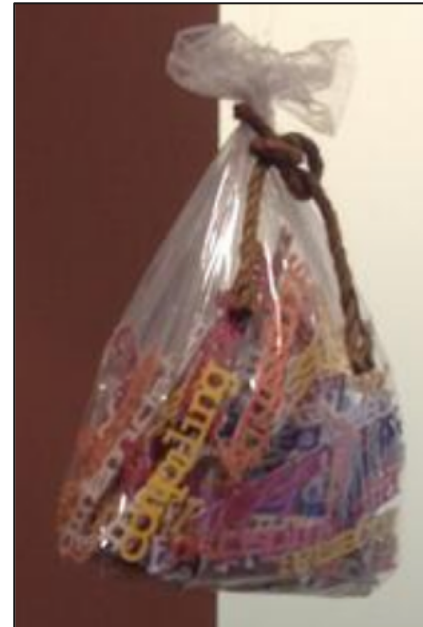
Jan-Mar



Apr-Jun



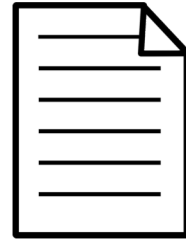
Jul-Sep



Oct-Dec



RQ2: Can we adapt to temporal variations?



Apr-Jun

General



Jan-Mar



Apr-Jun



Jul-Sep



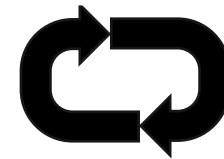
Oct-Dec



RQ2: Can we adapt to temporal variations?

- Straightforward to apply to seasonal features:

Data (Seasonal)	Baseline	Adaptation
Reviews (music)	.901	.919
Reviews (hotels)	.867	.881
Reviews (restaurants)	.874	.898
News (economy)	.782	.782
Twitter (vaccines)	.881	.880



RQ2: Can we adapt to temporal variations?

- How to use in non-seasonal settings?



2016

General



2012



2013



2014



2015



RQ2: Can we adapt to temporal variations?

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



2013

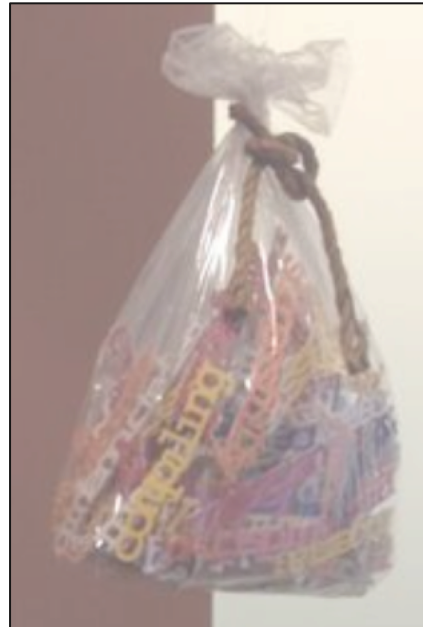
General

2012

2013

2014

2015



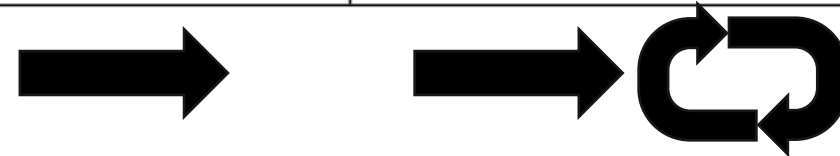
RQ2: Can we adapt to temporal variations?

- 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?

Data (Non-seasonal)	Baseline	Adaptation	Adapt.+seasons
Reviews (music)	.895	.924	.910
Reviews (hotels)	.886	.892	.920
Reviews (restaurants)	.831	.879	.889
News (economy)	.763	.780	.859
Politics (platforms)	.661	.665	n/a
Twitter (vaccines)	.910	.903	.920



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](https://github.com/xiaoleihuang/Domain_Adaptation_ACL2018)



College of Media, Communication
and Information
UNIVERSITY OF COLORADO **BOULDER**

