Adjusting for Confounding Using Text Matching

Molly Roberts (UCSD)

joint work with Brandon Stewart (Princeton) and Rich Nielsen (MIT)

April 29, 2021

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR
    - 3,000 Articles published in top tier IR 1980-2006

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR
  - 3,000 Articles published in top tier IR 1980-2006
  - Women receive about 80% of the citations of male counterparts

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR
  - 3,000 Articles published in top tier IR 1980-2006
  - Women receive about 80% of the citations of male counterparts
  - Network analysis suggests:

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR
  - 3,000 Articles published in top tier IR 1980-2006
  - Women receive about 80% of the citations of male counterparts
  - Network analysis suggests:
    - Women self-cite less than men

Does gender affect citations in Political Science?

- Lots of discussion about status of women in academia/implicit bias
- Maliniak, Powers, Walter (2013): women cited less than men in IR
    - 3,000 Articles published in top tier IR 1980-2006
    - Women receive about 80% of the citations of male counterparts
    - Network analysis suggests:
        - Women self-cite less than men
        - Men cite men more and men make up largest proportion of scholars

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
  - Randomly assign names to journal articles

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
    - Randomly assign names to journal articles
    - Wait for ten years

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
    - Randomly assign names to journal articles
    - Wait for ten years
    - See how many citations they accumulate over time

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
  - Randomly assign names to journal articles
  - Wait for ten years
  - See how many citations they accumulate over time
- Maliniak et al solution: TRIP codes articles into (many) categories

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
  - Randomly assign names to journal articles
  - Wait for ten years
  - See how many citations they accumulate over time
- Maliniak et al solution: TRIP codes articles into (many) categories
- Control for these categories (topic, approach, etc)

Does gender affect citations in Political Science?

- Difficult study to do, because there are lots of confounders
- For example: women write about different topics/with different words
- Is the reason they get fewer citations the choice of topics, or perceived gender?
- The perfect experiment:
  - Randomly assign names to journal articles
  - Wait for ten years
  - See how many citations they accumulate over time
- Maliniak et al solution: TRIP codes articles into (many) categories
- Control for these categories (topic, approach, etc)
- Also control for many other things (R1, tenure, co-author, journal, ..)

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
    1. Really time consuming

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
    1. Really time consuming
    2. Need to know the confounding categories ahead of time

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
    1. Really time consuming
    2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
    1. Really time consuming
    2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male
- Our plan:

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male
- Our plan: Find similar articles,

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male
- Our plan: Find similar articles, control for text and non-text features,

Hand Coding Vs. Automated Methods

- TRIP hand coding a heroic effort
- Two problems with hand coding:
  1. Really time consuming
  2. Need to know the confounding categories ahead of time
- Can we do this automatically?
- Can we estimate the ways the relevant topics to condition on?
- Our solution: Text analysis!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Merge with Maliniak et al data $\rightsquigarrow$ we can use their other variables, including gender, article age, tenure, etc.
- Compare All-Female to Co-ed/All-Male
- Our plan: Find similar articles, control for text and non-text features, estimate how gender influences citations.

Text Matching

Text Matching

- Text as pre-treatment confounder

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT conditioning on high-dimensional confounders is hard

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT conditioning on high-dimensional confounders is hard
  - You can't possibly condition on every word! (and you wouldn't want to)

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT conditioning on high-dimensional confounders is hard
  - You can't possibly condition on every word! (and you wouldn't want to)
  - We care about controlling for covariates predictive of treatment

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT conditioning on high-dimensional confounders is hard
  - You can't possibly condition on every word! (and you wouldn't want to)
  - We care about controlling for covariates predictive of treatment
  - But with text, we don't always know what predicts treatment

Text Matching

- Text as pre-treatment confounder $\rightsquigarrow$ a surprisingly frequent problem
- Applications
  - In International Relations, are women cited less frequently than men?
  - Does censorship change the behavior of social media users?
    - Naive approach: compare posts/posting rate after censorship of censored and uncensored users
    - Confounder: Text of what users wrote before censorship
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT conditioning on high-dimensional confounders is hard
  - You can't possibly condition on every word! (and you wouldn't want to)
  - We care about controlling for covariates predictive of treatment
  - But with text, we don't always know what predicts treatment

Was very little work on matching on high-dimensional confounders, now some great work (Mozer et al 2018, Veitch et al 2019, Keith et al 2020)

Text matching

Our approach:

1. Construct analogs to current methods
   - Propensity score matching $\rightsquigarrow$ Multinomial Inverse Regression
   - Coarsened exact matching $\rightsquigarrow$ Topically Coarsened Exact Matching
2. Identify benefits and drawbacks of each
3. Create a new method Topical Inverse Regression Matching (TIRM), by combining the two

Outline of the talk

- A quick review of matching for causal inference
- Text analogs to current matching methods
- Topical Inverse Regression Matching
- Applications

A quick review of matching

A quick review of matching

- Goal: estimate effect given conditional ignorability
  $t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$

A quick review of matching

- Goal: estimate effect given conditional ignorability
  $$t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$$
- Many approaches: propensity score matching, coarsened exact matching, genetic matching, synthetic matching, covariate-balanced propensity scores, entropy balancing, mahalanobis matching, optimal matching, full matching, matching frontier, ...

A quick review of matching

- Goal: estimate effect given conditional ignorability
  $t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$

- Many approaches: propensity score matching, coarsened exact matching, genetic matching, synthetic matching, covariate-balanced propensity scores, entropy balancing, mahalanobis matching, optimal matching, full matching, matching frontier, ...

- Today two of these strategies:

A quick review of matching

- Goal: estimate effect given conditional ignorability
  $t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$
- Many approaches: propensity score matching, coarsened exact matching, genetic matching, synthetic matching, covariate-balanced propensity scores, entropy balancing, mahalanobis matching, optimal matching, full matching, matching frontier, ...
- Today two of these strategies:
  1. model $p(t_i|\vec{x}_i) \rightsquigarrow$ propensity score matching (PSM)

A quick review of matching

- Goal: estimate effect given conditional ignorability
  $t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$
- Many approaches: propensity score matching, coarsened exact matching, genetic matching, synthetic matching, covariate-balanced propensity scores, entropy balancing, mahalanobis matching, optimal matching, full matching, matching frontier, ...
- Today two of these strategies:
  1. model $p(t_i|\vec{x}_i) \rightsquigarrow$ propensity score matching (PSM)
  2. match on all $\vec{x}_i \rightsquigarrow$ coarsened exact matching (CEM)

A quick review of matching

- Goal: estimate effect given conditional ignorability

  $$t_i \perp\!\!\!\perp y_i(1), y_i(0) | \vec{x}_i$$

- Many approaches: propensity score matching, coarsened exact matching, genetic matching, synthetic matching, covariate-balanced propensity scores, entropy balancing, mahalanobis matching, optimal matching, full matching, matching frontier, ...

- Today two of these strategies:
  1. model $p(t_i | \vec{x}_i) \rightsquigarrow$ propensity score matching (PSM)
  2. match on all $\vec{x}_i \rightsquigarrow$ coarsened exact matching (CEM)

- Both strategies scale poorly with high-dimensional covariates.

Getting propensity scores for text with MNIR

- Classical PSM approach:
  - fit logistic regression $\hat{\pi}_i = p(t_i|\vec{x}_i)$
  - match units with similar probability of treatment
  - pros: units matched by scalar ($\hat{\pi}_i$) instead of long vector ($\vec{x}_i$)
  - cons: approximates full randomization rather than more efficient block randomization (King and Nielsen 2019)

Getting propensity scores for text with MNIR

- Classical PSM approach:
  - fit logistic regression $\hat{\pi}_i = p(t_i|\vec{x}_i)$
  - match units with similar probability of treatment
  - pros: units matched by scalar ($\hat{\pi}_i$) instead of long vector ($\vec{x}_i$)
  - cons: approximates full randomization rather than more efficient block randomization (King and Nielsen 2019)

- Problem: high-dimensional confounders
  - $\boldsymbol{X}$ is $N \times V$ (# of documents by # of words in vocab)
  - can only estimate $\hat{\pi}_i$ well when $N \gg V$, which isn't the case!

Getting propensity scores for text with MNIR

- Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)
  - assume $x_i \sim$Multinomial$(\vec{q}_i, m_i = \sum_v x_{i,v})$
  - where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
  - ϕ_v measures relationship between treatment and word
  - projection $z_i = \Phi'(\vec{x}_i / m_i)$ is a sufficient reduction $\boldsymbol{X} \perp\!\!\!\perp T | Z$
    $\rightsquigarrow$ estimate $\hat{\pi}_i$ with projection
  - Match on z_i or $\hat{\pi}_i$

Getting propensity scores for text with MNIR

Problem:

Texts equally likely to be treated are not always semantically similar.
Wouldn't be a problem in expectation, but...

Getting propensity scores for text with MNIR

Problem:

Texts equally likely to be treated are not always semantically similar. Wouldn't be a problem in expectation, but...

- hard to assess balance in the text case
- could be more efficient if matches were more similar

Matching text with Coarsened Exact Matching analogs

- Classical CEM approach
  - coarsen each variable into natural categories
    i.e. years of education $\rightsquigarrow$ {high school, elementary school, college}
  - exactly match on coarsened variable
  - pros: bounds imbalance on each variable

Matching text with Coarsened Exact Matching analogs

- Classical CEM approach
  - coarsen each variable into natural categories
    i.e. years of education $\rightsquigarrow$ {high school, elementary school, college}
  - exactly match on coarsened variable
  - pros: bounds imbalance on each variable
- Problem: high-dimensional confounder set
  - thousands of variables, so no exact matches even if we coarsen

Matching text with Coarsened Exact Matching analogs

- Solution: topically coarsened matching
  - innovation: coarsen across variables
    simple example: "tax", "income", "tariff" $\rightsquigarrow$ "economics"
  - topics must be equivalent across documents instead of words
  - bounds imbalance across groups of stochastically equivalent words
- Estimate a topic model such as LDA (Blei, Ng and Jordan 2003)
- Match on the topic density rather than raw word counts
- Problem: topics aren't always import predictors of treatment.

Topical Inverse Regression Matching (TIRM)

We need something that:

1. Bounds imbalance between documents
2. Doesn't leave out important words

Topical Inverse Regression Matching (TIRM)

We need something that:

1. Bounds imbalance between documents
2. Doesn't leave out important words

Topical Inverse Regression Matching (TIRM)

- Jointly estimate probability of treatment and topic density
- Match on topic proportions & topic-specific probability of treatment
  - topical bounding properties
  - estimates which words associated with treatment
- Ingredients:
  - Structural Topic Model
  - with treatment as content covariate

Topic models

Two matrices estimated:

1) Topical Prevalence Matrix (DxK)

2) Topical Content Matrix (VxK)

Topic models

Two matrices estimated:

1) Topical Prevalence Matrix (DxK)

$$\theta = \left[\begin{array}{c|cccc} & Topic1 & Topic2 & \dots & TopicK \\ \hline Doc1 & .2 & .1 & \dots & 0.05 \\ Doc2 & .2 & .1 & \dots & .3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ DocD & 0 & 0 & \dots & .5 \end{array}\right]$$

2) Topical Content Matrix (VxK)

Topic models

Two matrices estimated:

1) Topical Prevalence Matrix (DxK)

$$\theta = \left[\begin{array}{c|cccc} & Topic1 & Topic2 & \ldots & TopicK \\ \hline Doc1 & .2 & .1 & \ldots & 0.05 \\ Doc2 & .2 & .1 & \ldots & .3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ DocD & 0 & 0 & \ldots & .5 \end{array}\right]$$

2) Topical Content Matrix (VxK)

$$\beta^{\boldsymbol{T}} = \left[\begin{array}{c|cccc} & Topic1 & Topic2 & \ldots & TopicK \\ \hline \text{``}text\text{''} & .02 & .001 & \ldots & 0.001 \\ \text{``}data\text{''} & .001 & .02 & \ldots & 0.001 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{``}analysis\text{''} & .01 & .01 & \ldots & 0.0005 \end{array}\right]$$

Topic models

Two matrices estimated: $X \approx \theta\beta$

1) Topical Prevalence Matrix (DxK)

$$\boldsymbol{\theta} = \begin{bmatrix} & \vline & Topic1 & Topic2 & \dots & TopicK \\ \hline Doc1 & \vline & .2 & .1 & \dots & 0.05 \\ Doc2 & \vline & .2 & .1 & \dots & .3 \\ \vdots & \vline & \vdots & \vdots & \ddots & \vdots \\ DocD & \vline & 0 & 0 & \dots & .5 \end{bmatrix}$$

2) Topical Content Matrix (VxK)

$$\boldsymbol{\beta^T} = \begin{bmatrix} & \vline & Topic1 & Topic2 & \dots & TopicK \\ \hline \text{“}text'' & \vline & .02 & .001 & \dots & 0.001 \\ \text{“}data'' & \vline & .001 & .02 & \dots & 0.001 \\ \vdots & \vline & \vdots & \vdots & \ddots & \vdots \\ \text{“}analysis'' & \vline & .01 & .01 & \dots & 0.0005 \end{bmatrix}$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$\boldsymbol{\theta}_i | \boldsymbol{\alpha} \sim \text{Dirichlet}(\boldsymbol{\alpha})$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$\begin{aligned} \boldsymbol{\theta}_i | \boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\ \boldsymbol{z}_{im} | \boldsymbol{\theta}_i &\sim \text{Multinomial}(1, \boldsymbol{\theta}_i) \end{aligned}$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$\begin{aligned}
\boldsymbol{\theta}_i|\boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
\boldsymbol{z}_{im}|\boldsymbol{\theta}_i &\sim \text{Multinomial}(1, \boldsymbol{\theta}_i) \\
w_{im}|\boldsymbol{\beta}_k, z_{imk} = 1 &\sim \text{Multinomial}(1, \boldsymbol{\beta}_k)
\end{aligned}$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$\boldsymbol{\beta}_k \sim \text{Dirichlet}(\mathbf{1})$$

$$\begin{aligned}
\boldsymbol{\theta}_i|\boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
\boldsymbol{z}_{im}|\boldsymbol{\theta}_i &\sim \text{Multinomial}(1, \boldsymbol{\theta}_i) \\
w_{im}|\boldsymbol{\beta}_k, z_{imk} = 1 &\sim \text{Multinomial}(1, \boldsymbol{\beta}_k)
\end{aligned}$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \ldots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$
\begin{aligned}
\boldsymbol{\beta}_k &\sim \text{Dirichlet}(\mathbf{1}) \\
\alpha_k &\sim \text{Gamma}(\alpha, \beta) \\
\boldsymbol{\theta}_i | \boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
\boldsymbol{z}_{im} | \boldsymbol{\theta}_i &\sim \text{Multinomial}(1, \boldsymbol{\theta}_i) \\
w_{im} | \boldsymbol{\beta}_k, z_{imk} = 1 &\sim \text{Multinomial}(1, \boldsymbol{\beta}_k)
\end{aligned}
$$

Topic models

For Latent Dirichlet Allocation...

- Consider document i, $(i = 1, 2, \dots, D)$.
- Suppose there are M_i total words and $\boldsymbol{w}_i$ is an $M_i \times 1$ vector, where w_{im} describes the m^{th} word used in the document.

$$
\begin{aligned}
\boldsymbol{\beta}_k &\sim \text{Dirichlet}(\mathbf{1}) \\
\alpha_k &\sim \text{Gamma}(\alpha, \beta) \\
\boldsymbol{\theta}_i | \boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
\boldsymbol{z}_{im} | \boldsymbol{\theta}_i &\sim \text{Multinomial}(1, \boldsymbol{\theta}_i) \\
w_{im} | \boldsymbol{\beta}_k, z_{imk} = 1 &\sim \text{Multinomial}(1, \boldsymbol{\beta}_k)
\end{aligned}
$$

Optimize with Variational Inference or Gibbs Sampling.

Structural Topic Model

- Adds “structure” to LDA via a prior

  (Blei and Lafferty 2006, Mimno and McCallum 2008)
- Documents have different expected topic proportions based on observed covariates.
- Topics are now deviations from a baseline distribution.

Structural Topic Model

- Adds "structure" to LDA via a prior

  (Blei and Lafferty 2006, Mimno and McCallum 2008)
- Documents have different expected topic proportions based on observed covariates.
- Topics are now deviations from a baseline distribution.

$$P(word|topic, doc) \propto$$

$$\exp(\kappa^{(m)} + \text{topic} * \kappa^{(k)} + \text{covariate}_{doc} * \kappa^{(c)} + \text{topic*covariate}_{doc} * \kappa^{(int)})$$

Structural Topic Model

- Adds "structure" to LDA via a prior

  (Blei and Lafferty 2006, Mimno and McCallum 2008)
- Documents have different expected topic proportions based on observed covariates.
- Topics are now deviations from a baseline distribution.

$P(word|topic, doc) \propto$

$$\exp(\kappa^{(m)} + \text{topic} * \kappa^{(k)} + \text{covariate}_{doc} * \kappa^{(c)} + \text{topic*covariate}_{doc} * \kappa^{(int)})$$

$\kappa^{(c)}$ and $\kappa^{(int)} \rightsquigarrow$ how words are related to treatment.

Topical Inverse Regression Matching (TIRM)

First: Re-estimate θ as though document was treated.

Match on:

1. $\boldsymbol{\theta}$: Estimated topic proportion (K covariates)
2. ***projection***:
   - let (x_i/m_i) % of document i that is word x
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection
   - $(\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i}\sum_v x_{i,v}\left(\left(\kappa_v^{(\text{int})}\right)'\theta_i\right)$ topic-covariate projection
3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Topical Inverse Regression Matching (TIRM)

First: Re-estimate θ as though document was treated.

Match on:

1. $\boldsymbol{\theta}$: Estimated topic proportion (K covariates)
2. ***projection***:
   - let (x_i/m_i) % of document i that is word x
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection
   - $(\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i}\sum_v x_{i,v}\left(\left(\kappa_v^{(\text{int})}\right)'\theta_i\right)$ topic-covariate projection
3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Limitations of TIRM

- The regular... requires SUTVA, relevant covariates
- plus... relies on a parametric method to reduce dimensions

Balance metrics

No single unified balance metric, so we have to use a few:

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
    - use mutual information, MNIR, STM to estimate words close to treatment

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words
  - allows us to ensure bag of words is representing text

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words
  - allows us to ensure bag of words is representing text
- Human comparisons

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words
  - allows us to ensure bag of words is representing text
- Human comparisons (easier for Tweets hard for novels)

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words
  - allows us to ensure bag of words is representing text
- Human comparisons (easier for Tweets hard for novels)
  - Human coded categories

Balance metrics

No single unified balance metric, so we have to use a few:

- Balance on words associated with treatment
  - use mutual information, MNIR, STM to estimate words close to treatment
  - try to achieve balance on words associated with treatment post matching
- Balance on estimated topics
- String kernel similarity (Lohdi et al 2002, Spirling 2012)
  - measures similarity in sequences of words
  - allows us to ensure bag of words is representing text
- Human comparisons (easier for Tweets hard for novels)
  - Human coded categories
  - User reads sample of paired matches, assesses similarity

Application: Gender bias in citations

Setting

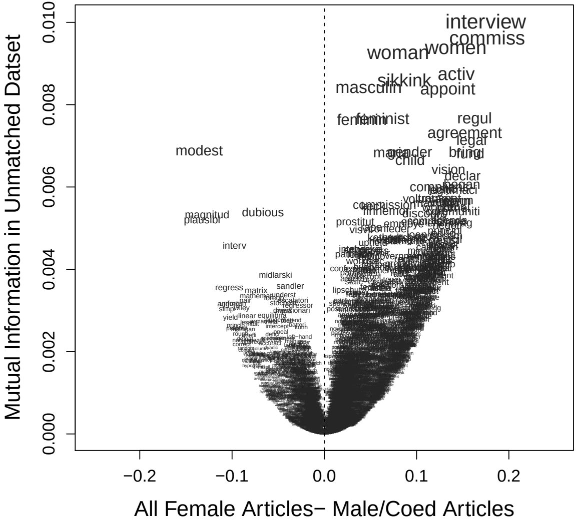
- Maliniak, Powers, Walter (2013): women get cited less than men in political science
- ...but women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching

Data: 3,201 journal articles from top 12 IR journals, 1980-2006.

- Lots of variables, including gender, article age, tenure, etc.
- Treatment: all-female vs. control: co-ed/all-male
- Goal: Find similar articles, see how they are cited differently.

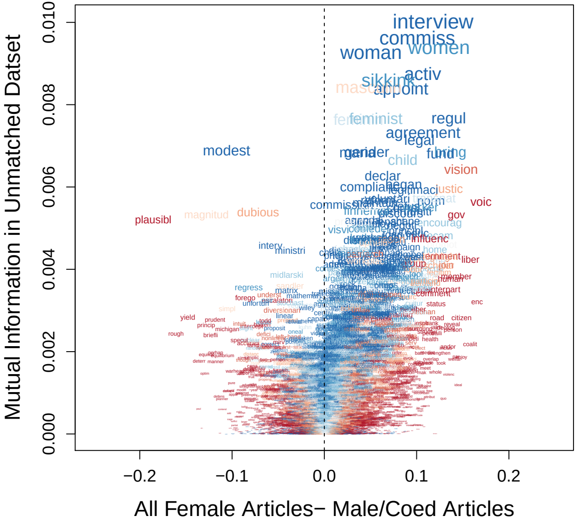
Application: Gender bias in citations

Original data:
No matching



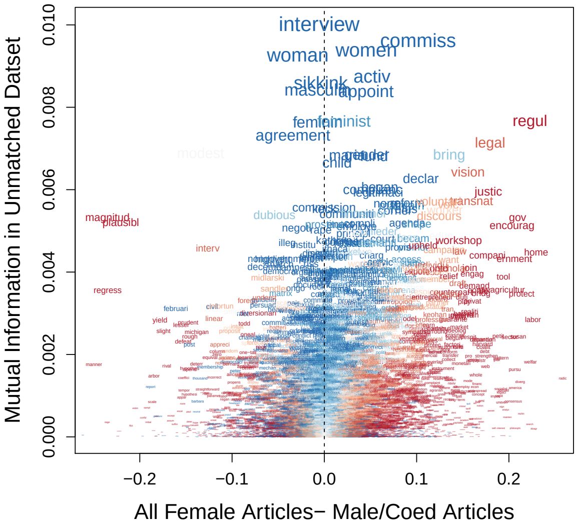
Application: Gender bias in citations

Matched data:
Topical CEM

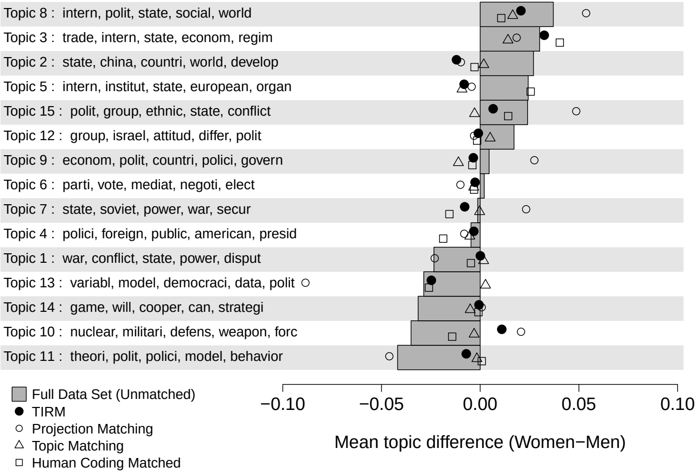


Application: Gender bias in citations

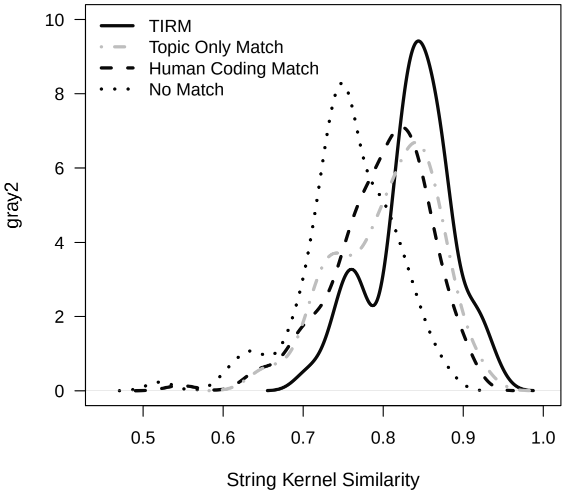
Matched data:
TIRM



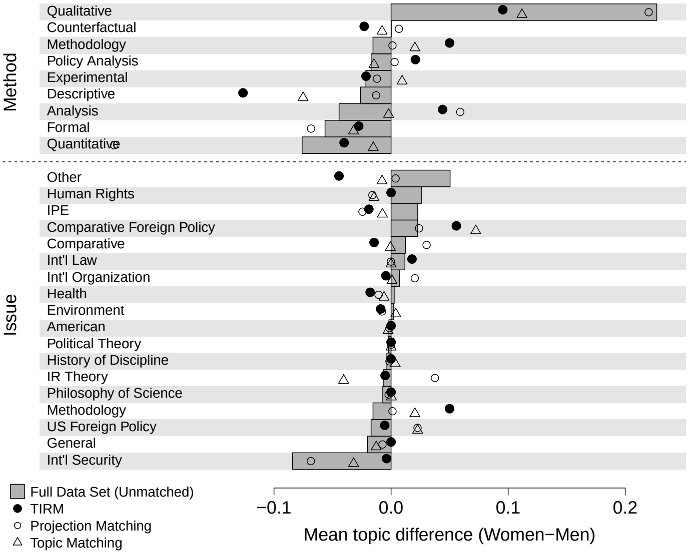
TIRM Reduces Topical Differences



TIRM improves string kernel similarity



TIRM Reduces Human Coded Differences



Application: Gender bias in citations

Results

- Maliniak et al: Women receive 80% of the citations of men
- We find: women receive fewer citations (robust across specifications)
- Our estimate: Women receive 65% of the citations of men
- The difference is in very high expected citation counts:
  - Low range: 14 cites vs. 12 cites, not statistically detectable diff.
  - High range: 90 cites vs. 20 cites, very easy to detect.

Conclusion

Conclusion

- Lots of applications measure pre-treatment confounders with text

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics
  - Matching on probability of treatment

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics
  - Matching on probability of treatment $\rightsquigarrow$ balances on words related to treatment

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics
  - Matching on probability of treatment $\rightsquigarrow$ balances on words related to treatment
- is best for overall balance.

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics
  - Matching on probability of treatment $\rightsquigarrow$ balances on words related to treatment
- is best for overall balance.
- Future work:

Conclusion

- Lots of applications measure pre-treatment confounders with text
- We propose a new framework and show that:
  - Matching on topical density estimate $\rightsquigarrow$ bounds differences between topics
  - Matching on probability of treatment $\rightsquigarrow$ balances on words related to treatment
- is best for overall balance.
- Future work:
  - Extend to high-dimensional cases other than text