More scaling

William Lowe

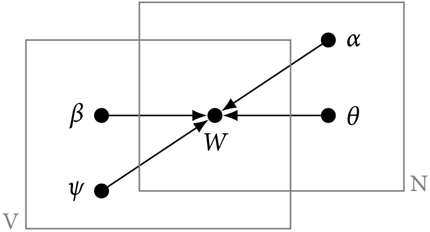
Hertie School

19th November 2020

Plan

→ Model fitting
→ Interpreting the dimension
→ Specifying the dimension
→ Making things comparable
→ Going multidimensional
→ Visualisation
→ Searching for a common space

The model (ML version)



$$\log C_{ij} = \alpha_i + \psi_j + \theta_i \beta_j$$

Model fitting by 'coordinate ascent' (Goodman, 1979)

0. Guess θ and α
1. Fit each β_j (as slope) and ψ_j (as intercept) as a Poisson regression with offset α and 'covariates' θ
2. Fit each θ_i (as slope) and α_i (as intercept) in a Poisson regression with offset ψ and 'covariates' β
3. If the likelihood is not yet maximized, go to step 2.

Slow, but reasonably reliable.

Unidimensionality

What is this θ anyway?

- → Whatever fits the data best. You have to interpret them

Last week's 'spatial talking' implies that

- → Document positions are an average of the βs of the words in them (and word positions are an average of the documents they appear in)
- → So if we can interpret the βs as a substantive scale, we know that the dimension is

Unidimensionality

What is this θ anyway?

- → Whatever fits the data best. You have to interpret them

Last week's 'spatial talking' implies that

- → Document positions are an average of the βs of the words in them (and word positions are an average of the documents they appear in)
- → So if we can interpret the βs as a substantive scale, we know that the dimension is

Not everything has to be a dimension

- → but it does for a scaling model!

Difficult cases:

- → Populism and anti-system parties. Are they well understood as ideological?
- → Government and opposition. Naturally polar but not necessarily ideologically so

Unidimensionality

Related questions

→ How do we know that positions are only one dimension?

→ How to get positions on a *specific* policy issue?

Unidimensionality

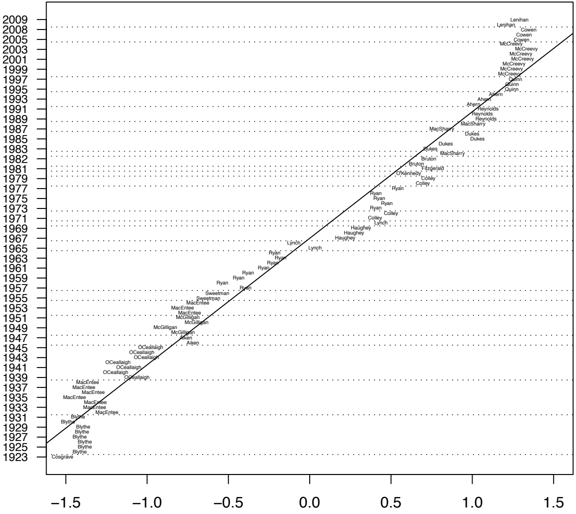
Related questions

→ How do we know that positions are only one dimension?
→ How to get positions on a *specific* policy issue?

Possibilities, in order of directness

→ Select vocabulary that is on topic (e.g. in sentiment analysis; Proksch et al., 2019)
→ Use topic counts instead of words (e.g. RILE; Budge et al., 1987)
→ Scale only on-topic segments of the each document (Slapin & Proksch, 2008)
→ Learn β from (a subset of) known 'reference' documents ('Wordscores'; Laver et al., 2003)
→ Trust institutional or strategic context to constrain to one dimension (Baerg & Lowe, 2020)

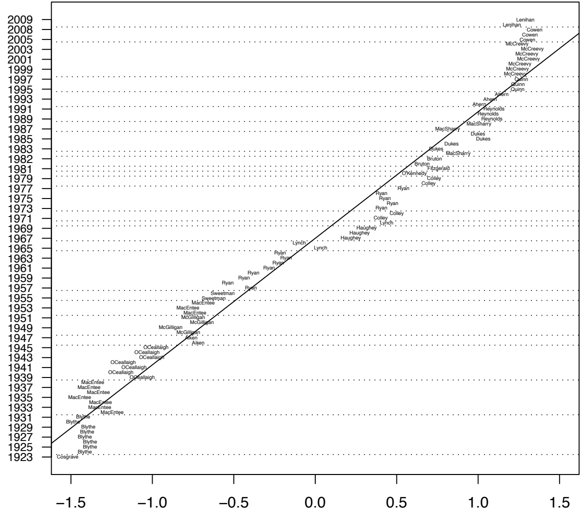
Oops



All the budget speeches in independent Irish history, scaled.

- → Budgets are about spending money on things
- → Those things change over time
- → The model cannot know

Oops



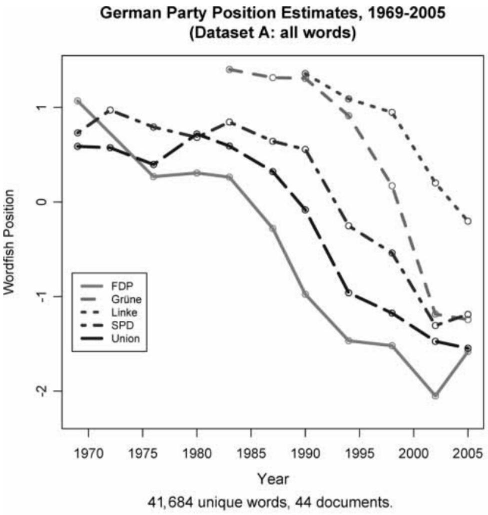
All the budget speeches in independent Irish history, scaled.

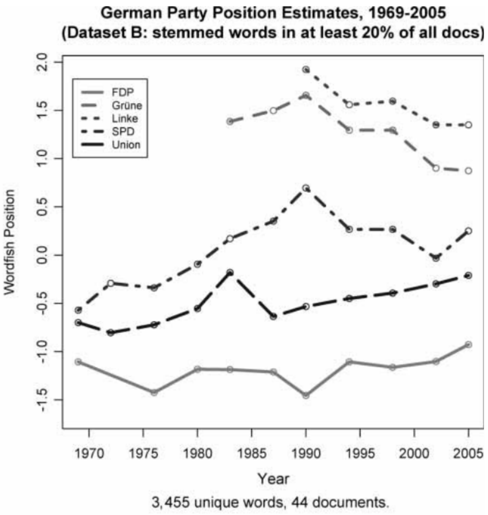
- → Budgets are about spending money on things
- → Those things change over time
- → The model cannot know

Possible fix

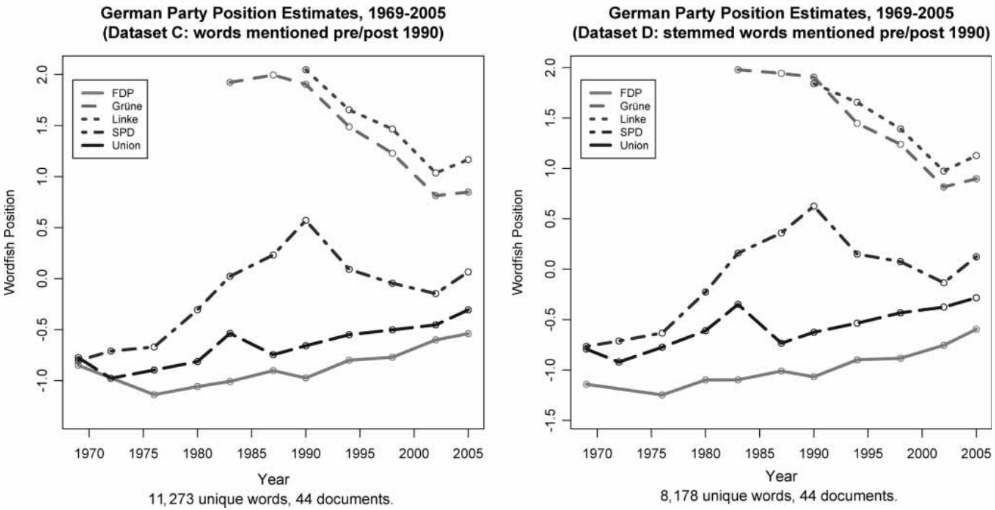
- → Choose a subset of stable vocabulary (Proksch & Slapin, 2009)

Vocabulary decisions





Vocabulary decisions



Wordscores

Wordscores is an early approach to topic-specific unidimensional scaling (Laver et al., 2003)

0. Assert the positions of R reference documents $\theta_1 \ldots \theta_R$
1. Using only these, estimate β (there are no ψs)
2. Estimate each remaining θs as the average of the βs of its words

$$\hat{\theta}_i = \frac{1}{N}\sum_i^N \beta_{v[i]} = \sum_v^V \beta_v F(v \mid \text{document } i)$$

Wordscores

Wordscores is an early approach to topic-specific unidimensional scaling (Laver et al., 2003)

0. Assert the positions of R reference documents $\theta_1 \ldots \theta_R$
1. Using only these, estimate β (there are no ψs)
2. Estimate each remaining θs as the average of the βs of its words

$$\hat{\theta}_i = \frac{1}{N}\sum_i^N \beta_{v[i]} = \sum_v^V \beta_v F(v \mid \text{document } i)$$

Why does this work (when it does work?)

Intuition:

→ Variation across $\theta_1 \ldots \theta_R$ defines the dimension
→ When word v varies a lot over documents with similar θs then β_v must be small because it shouldn't matter to the doesn't to $\hat{\theta}$
→ …Profit?

The model (geometrical version)

We can also a lot of dimensions at once

- → Not straightforward with maximum likelihood
- → Very easy if we move to least squares

Linear probability model:

- → Previously we modeled (the log of) $E[C_{ij}]$
- → This time we will model the proportion of corpus's counts that are document i using word j

$$P_{ij} = C_{ij}/N$$

If we compute the marginal proportions

$$P_i = \sum_j P_{ij} \qquad\qquad P_j = \sum_i P_{ij}$$

then the probability of seeing word j in document i 'by chance' is $P_i P_j$ (the 'independence model')

The model (geometrical version)

Define a normalized *residual*

$$R_{ij} = \frac{P_{ij} - P_i P_j}{\sqrt{P_i P_j}}$$

decompose it by SVD to get orthogonal 'positions'

$$R = U \Sigma V^T$$

Identify the first k columns of U as $\theta_1 \ldots \theta_K$ and the top k columns of V as $\beta_1 \ldots \beta_K$

The model (geometrical version)

Define a normalized *residual*

$$R_{ij} = \frac{P_{ij} - P_i P_j}{\sqrt{P_i P_j}}$$

decompose it by SVD to get orthogonal 'positions'

$$R = U \Sigma V^T$$

Identify the first k columns of U as $\theta_1 \ldots \theta_K$ and the top k columns of V as $\beta_1 \ldots \beta_K$

$$P_{ij} \approx P_i P_j \left(1 + \sum_k^K \theta_i^{(k)} \sigma^{(k)} \beta_j^{(k)}\right)$$

This is *correspondence analysis* (Benzécri, 1992; Greenacre, 2007)

Compare and contrast

This tends to be very close to the ML version of out model

- → But we get K dimensional positions
- → But no standard errors, since it's a geometrical model

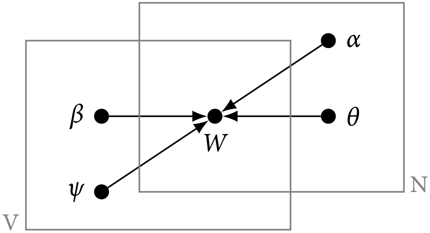
Why is it similar? Because

$$\log\left(1 + \sum_k^K \theta_i^{(k)} \sigma^{(k)} \beta_j^{(k)}\right) \approx \sum^K \theta_i^{(k)} \sigma^{(k)} \beta_j^{(k)}$$

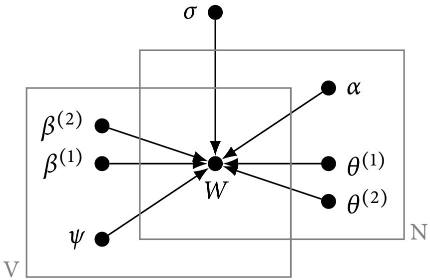
when there there is smallish amounts of positioning happening.[1]

[1]Note: the left hand side parameters are from CA and the right hand side are from Association/Wordfish model. I just haven't distinguished them in the notation because I am a bad person.

Multidimensional positions

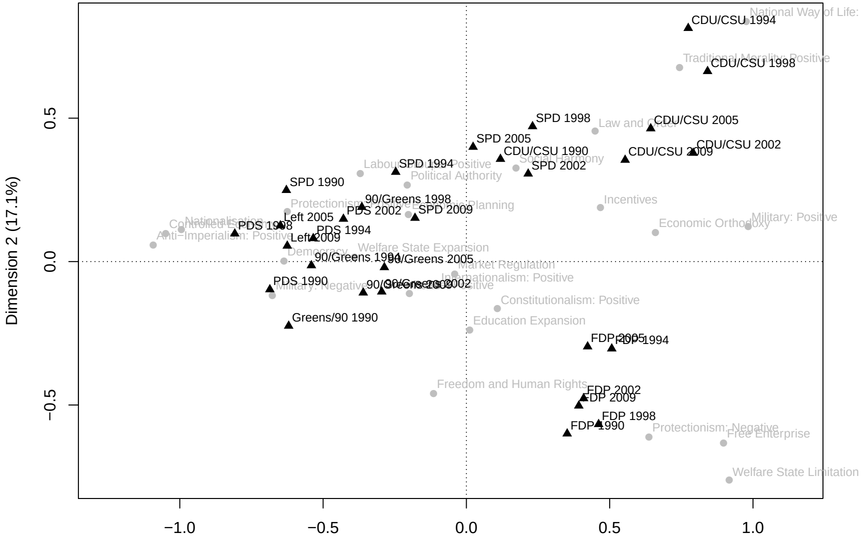


$$\log C_{ij} = \alpha_i + \psi_j + \theta_i \beta_j$$



$$\log C_{ij} = \alpha_i + \psi_j + \sum_k \theta_i^{(k)} \sigma^{(k)} \beta_j^{(k)}$$

Partying in 2 dimensions

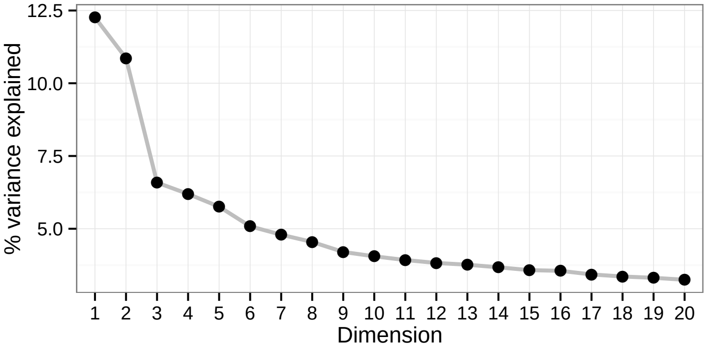


Life skills

How to read a biplot:

→ Documents points are closer when using words/topics similarly
→ Words points are closer with *similar* document profiles
→ 0,0: a document or word/topic used exactly as often as we would expect by chance
→ Document vector: arrow from 0,0 to a document point
→ Word/topic vector: arrow from 0,0 to a word/topic point
→ Vectors are *longer* the more their usage diverges from chance
→ Angle between a word vector and document vector: how much a document preferentially uses the word

Why was that a two dimensional plot?



How much variance is explained by adding another dimension?

→ Remember those σs?

What are these dimensions?

Dimensions are necessarily orthogonal

- → We expect substantive dimensions to be more or less correlated
- → Unlike e.g. factor analysis we are conditionally Poisson / Multinomial, so there is no separate correlation parameter
- → Lack of rotation keeps us honest…

At most, only 'intrinsic dimensionality' is really meaningful

- → Makes sense for stable party systems, with notable exceptions
- → May make less sense for individuals (Broockman, 2016)

Common space

What we want

→ A ’common space’ (Klüver, 2009)

How to get one

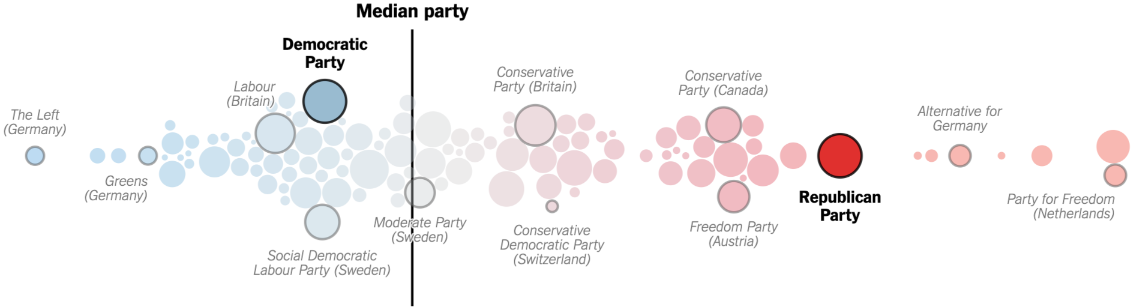
→ fiat (the traditional answer) and projection
→ bridging observations

Differential item functioning (DIF) may stop us

→ across space: does ‘Anti Imperialism’ really mean the same things in the West, the former East, and Central America?
→ across time: pre and post 1989 in Europe

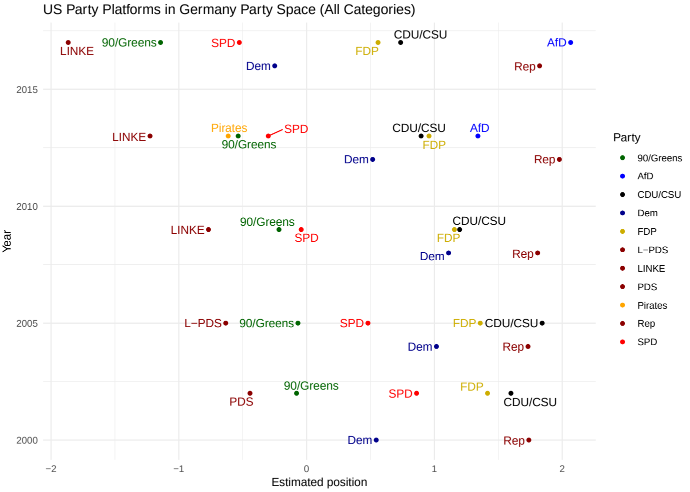
These are just different ways of talking about institutional stability

Common space: big stretch



Note: Circles sized by the percentage of the vote won by the party in the latest election in this data. Only parties that won more than 1 percent of the vote and are still in existence are shown. We analyzed parties in a selection of Western European countries, Canada and the United States.

Common space: smaller stretch?



Summing up

We can

- → Fit mutiple unidimensional models
- → Fit multidimensional models
- → Project new vectors of counts into existing spaces (βs)
- → Visualise pair of dimensions using biplots

But in the end, we have to make sense of it all ourselves…

References

Baerg, N. & Lowe, W. (2020). 'A textual taylor rule: Estimating central bank preferences combining topic and scaling methods'. *Political Science Research and Methods*, *8*(1), 106–122.

Benzécri, J.-P. (1992). 'Correspondence analysis handbook'. Marcel Dekker.

Broockman, D. E. (2016). 'Approaches to studying policy representation'. *Legislative Studies Quarterly*, *41*(1), 181–215.

Budge, I., Robertson, D. & Hearl, D. (Eds.). (1987). 'Ideology, strategy and party change: Spatial analyses of post-war election programmes in 19 democracies'. Cambridge University Press.

Goodman, L. A. (1979). 'Simple models for the analysis of association in cross-classifications having ordered categories'. *Journal of the American Statistical Association*, *74*(367), 537–552.

References

Greenacre, M. (2007). 'Correspondence analysis in practice'. Chapman; Hall/CRC.

Klüver, H. (2009). 'Measuring interest group influence using quantitative text analysis'. *European Union Politics*, *10*(4), 535–549.

Laver, M., Benoit, K. R. & Garry, J. (2003). 'Extracting policy positions from political texts using words as data'. *American Political Science Review*, *97*(2), 311–331.

Proksch, S.-O., Lowe, W., Wäckerle, J. & Soroka, S. (2019). 'Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches'. *Legislative Studies Quarterly*, *44*(1), 97–131.

Proksch, S.-O. & Slapin, J. B. (2009). 'How to avoid pitfalls in statistical analysis of political texts: The case of germany'. *German Politics*, *18*(3), 323–344.

Slapin, J. B. & Proksch, S.-O. (2008). 'A scaling model for estimating time-series party positions from texts'. *American Journal of Political Science*, *52*(3), 705–722.