WHOSE SCALE IS IT ANYWAY?

Will Lowe

Hertie School

October 20, 2021

'SCHLAND!

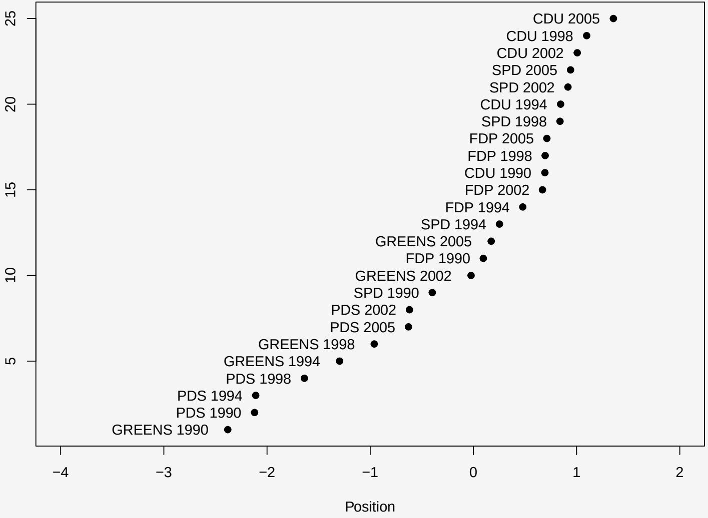
German party positions, estimated from their *platform texts* (Slapin & Proksch, 2008)

Not shown here:

→ positions of about 9000 words that place these parties where they are

Cases and items…

In one dimension…(and no time series either)

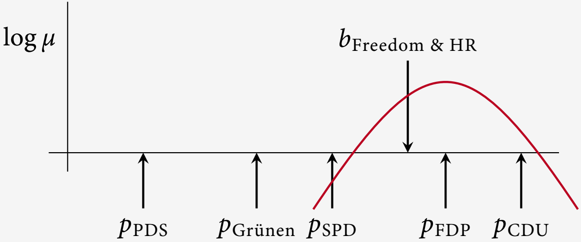


QUESTIONS ABOUT TEXT SCALING

- → What *positioning logic* do scaling models realize?
- → How do we place documents in *multiple dimensions*?
- → How do we *interpret* the results when we do?
- → How do we control the *substance* of a scale?

I mean, what could possibly go wrong?

THE LOGIC OF POSITION



$$
\begin{aligned}
\log \mu_{iF} &= r_i + c_F - \frac{1}{2}\frac{(p_i - b_F)^2}{\nu} \\
&= [r_i - p_i^2/\nu] + [c_F - b_F^2/\nu] + p_i\,[1/\nu]\,b_F \\
&= \alpha_i + \psi_F + \theta_i \sigma \beta_F
\end{aligned}
$$

Wordfish (Slapin & Proksch, 2008) is the *reduced form* of a spatial choice model with quadratic utility (Lowe, 2016)

→ Clinton et al. (2004) is spatial voting

→ Wordfish is spatial talking (e.g. Elff, 2013)

MOAR POSITIONS

Multidimensional extensions are conceptually straightforward: Just add more positions

$$\log \mu_{iF} = \alpha_i + \psi_F + \sum_m^M \theta_i^m \sigma^m \beta_F^m$$

This is the RC(M) association model (Goodman, 1979, 1981)

MOAR POSITIONS

Multidimensional extensions are conceptually straightforward: Just add more positions

$$\log \mu_{iF} = \alpha_i + \psi_F + \sum_m^M \theta_i^m \sigma^m \beta_F^m$$

This is the RC(M) association model (Goodman, 1979, 1981)

Note

Positions are parameters and counts are Poisson (or Multinomial, conditional on document length), so

→ dimensions must be *orthogonal* for identification

→ unlike everything we believe about policy dimensions...

Compare and contrast

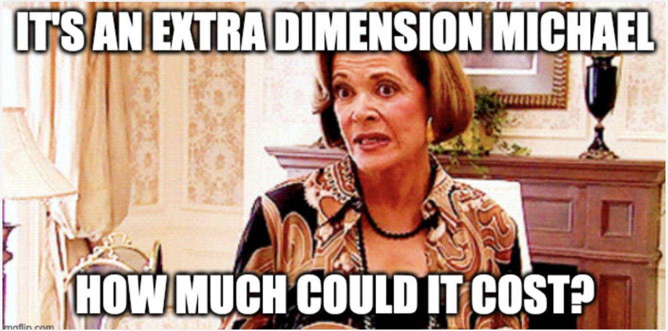
Factor analysis assumes

→ 'dominance' not ideal point measurement structure

→ responses are conditionally Normal, so θ is *rotatable* in all kinds of [ahem] exciting ways

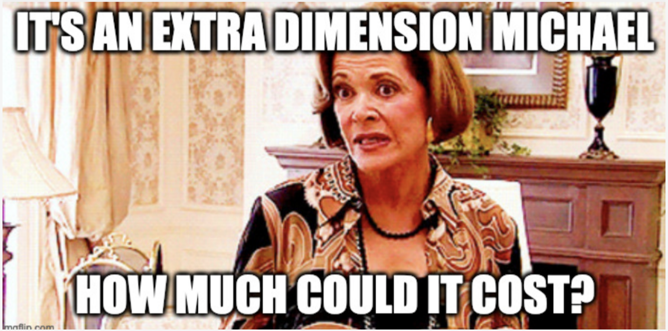
MOAR POSITIONS

These models are not widely used…



MOAR POSITIONS

These models are not widely used…



because they are *excruciating hard to fit* (and slightly unstable) especially in high dimensions

→ See {gnm} or {logmult} (Turner & Firth, 2008) if this sounds like your idea of a good time

MOAR POSITIONS

One solution is the approximation

$$P_{iF} \;\approx\; P_i P_F \left(1 + \sum_m^M \tilde{\theta}_i^m \tilde{\sigma}^m \tilde{\beta}_F^m\right)$$

a.k.a. correspondence analysis (Benzécri, 1992; Greenacre, 2007)

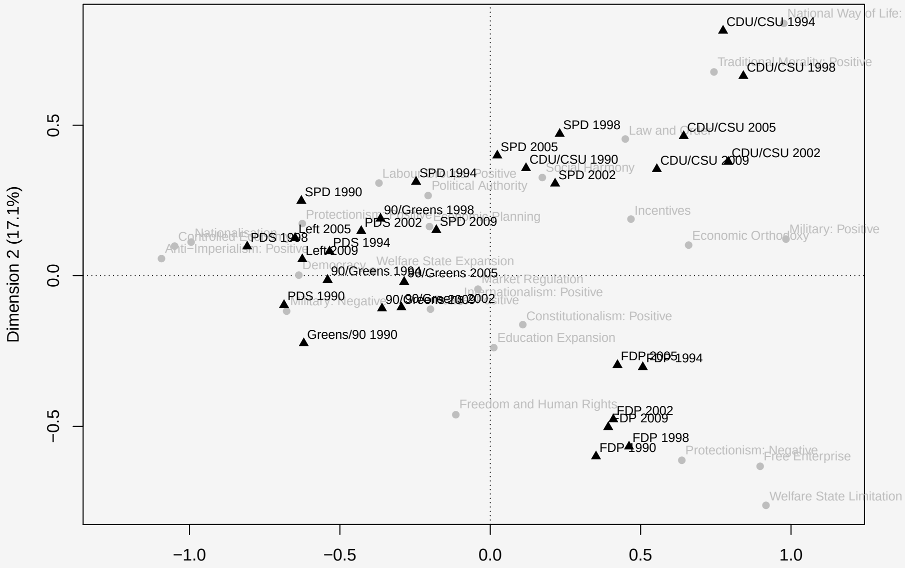
Price: One SVD

When $\sum_m^M \tilde{\theta}_i^m \tilde{\sigma}^m \tilde{\beta}_F^m$ is small

$$\exp\left(\sum_m^M \theta_i^m \sigma^m \beta_F^m\right) \;\approx\; 1 + \sum_m^M \tilde{\theta}_i^m \tilde{\sigma}^m \tilde{\beta}_F^m$$

Goodman (1981) compares the two in detail.

MULTI! MULTI! MULTI!



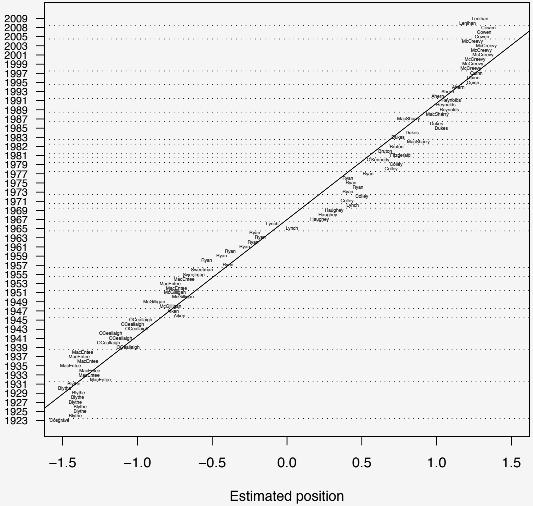
INTERPRETING MULTIDIMENSIONAL SCALING MODELS

INTERPRETING MULTIDIMENSIONAL SCALING MODELS

→ Familiar trouble

→ Unfamiliar trouble

FAST AND NUMEROUS: DUBLIN DRIFT



Exploratory multidimensional scaling models are like a box of chocolates; You never know what you're going to get

Intuition:

- → Items name core budget items + year-specific items
- → Ideological positioning can only happens *within* each budget's item set

Item turnover is what the ecologists refer to as a 'gradient length' (Legendre & Legendre, 2003)

THE PLATONIC FORM OF UNIDIMENSIONAL DATA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1	2	4	7	8	7	4	2	1	.	.	.	.	.	.	.	.	.	.
2	.	1	2	4	7	8	7	4	2	1	.	.	.	.	.	.	.	.	.
3	.	.	1	2	4	7	8	7	4	2	1	.	.	.	.	.	.	.	.
4	.	.	.	1	2	4	7	8	7	4	2	1	.	.	.	.	.	.	.
5	.	.	.	.	1	2	4	7	8	7	4	2	1	.	.	.	.	.	.
6	.	.	.	.	.	1	2	4	7	8	7	4	2	1	.	.	.	.	.
7	.	.	.	.	.	.	1	2	4	7	8	7	4	2	1	.	.	.	.
8	.	.	.	.	.	.	.	1	2	4	7	8	7	4	2	1	.	.	.
9	.	.	.	.	.	.	.	.	1	2	4	7	8	7	4	2	1	.	.
10	.	.	.	.	.	.	.	.	.	1	2	4	7	8	7	4	2	1	.
11	.	.	.	.	.	.	.	.	.	.	1	2	4	7	8	7	4	2	1

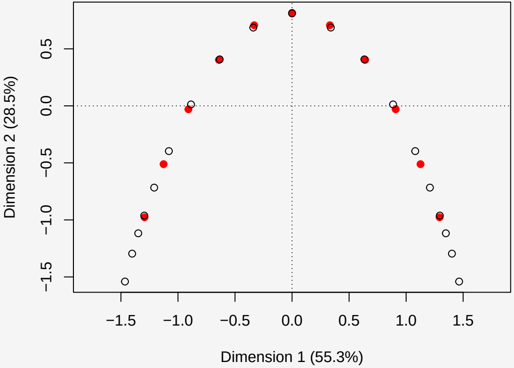
A long gradient (minus Poisson noise). If we don’t know, we might fit a multidimensional model…

BACK IN THE CAVE

Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001

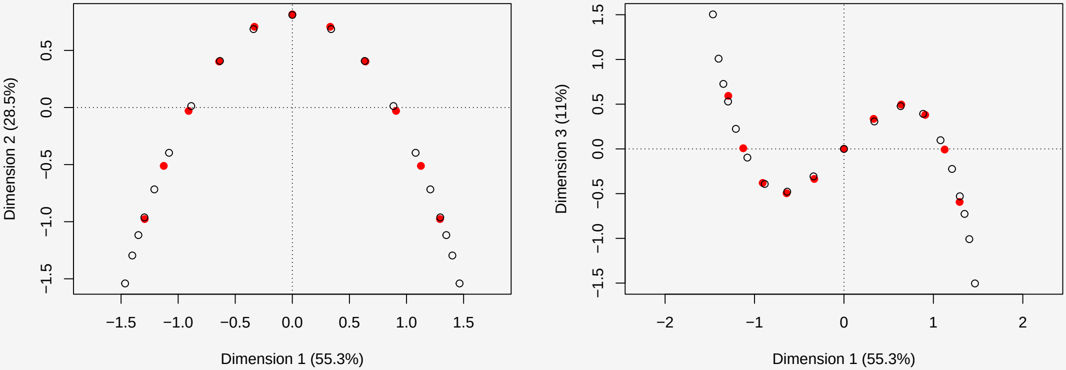
BACK IN THE CAVE

Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001



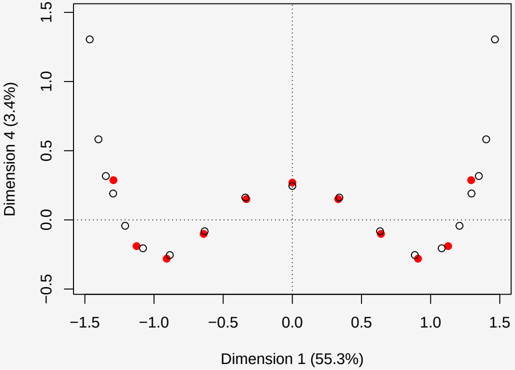
BACK IN THE CAVE

Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001



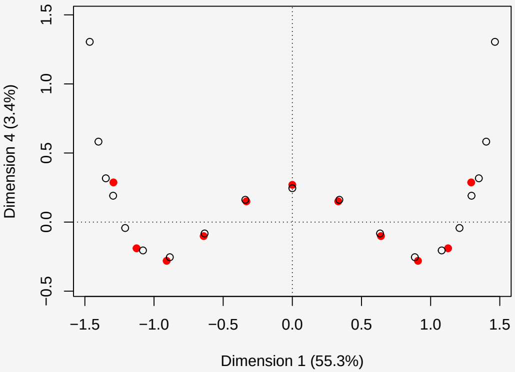
BACK IN THE CAVE

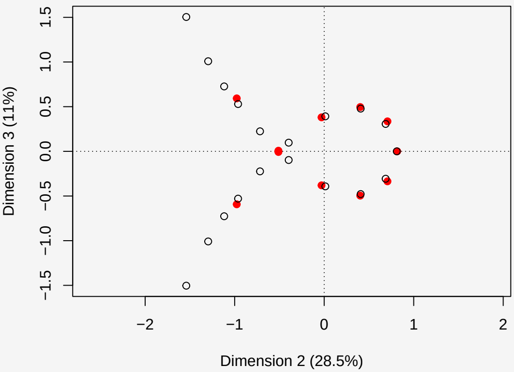
Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001



BACK IN THE CAVE

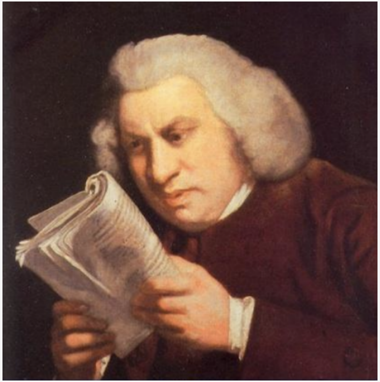
Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001





THE WHAT?

Dim.	1	2	3	4	5	6	7	8	9	10
Var.	0.553	0.285	0.11	0.034	0.01	0.004	0.002	0.001	<0.001	<0.001



THE ARCH

This is the *arch* (or horseshoe, or 'Guttman effect')

INTUITION (such as it is)

→ Scaling models are low dimensional reconstructions of *distance* (χ^2 distance)

→ Documents are distant when they have different word / topic frequencies

→ Words / topics are distant when they have different document profiles

→ Document 1 and 11 don't share any words

THE ARCH

This is the *arch* (or horseshoe, or 'Guttman effect')

MORE INTUITION (such as it is)

→ Folding the first dimension in half optimises inter-document distance under orthogonality constraints…

Caveats:

→ The arch is artifactual when there is *no* multidimensional structure

→ When there is, we'll get a mixture of it and the 'real' structure

THE ARCH

This is the *arch* (or horseshoe, or 'Guttman effect')

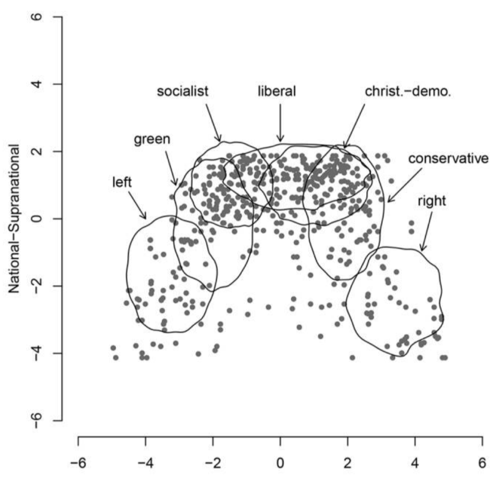
MORE INTUITION (such as it is)

→ Folding the first dimension in half optimises inter-document distance under orthogonality constraints…

Caveats:

→ The arch is artifactual when there is *no* multidimensional structure

→ When there is, we'll get a mixture of it and the 'real' structure



(König et al., 2017)

ARCHERY FOR BEGINNERS

'Effect' size is a function of overlap. We'll measure it with ideal point separation (here 6)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	2	4	7	**8**	7	4	2	1	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	1	2	4	7	**8**	7	4	2	1	.	.	.	.

ARCHERY FOR BEGINNERS

‘Effect’ size is a function of overlap. We’ll measure it with ideal point separation (here 6)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	2	4	7	**8**	7	4	2	1	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	1	2	4	7	**8**	7	4	2	1	.	.	.	.

	Dim.	1	2	3	4	5	6	7	8	9	10
Var. expl.	1	0.553	0.285	0.110	0.034	0.010	0.004	0.002	0.001	0.001	0.000
	2	0.306	0.253	0.185	0.120	0.070	0.036	0.017	0.007	0.003	0.002
	3	0.207	0.190	0.165	0.135	0.104	0.076	0.052	0.034	0.021	0.014
	4	0.158	0.151	0.139	0.124	0.108	0.091	0.075	0.061	0.050	0.043
	5	0.131	0.127	0.121	0.114	0.105	0.095	0.086	0.079	0.073	0.069
	6	0.116	0.114	0.111	0.107	0.102	0.098	0.093	0.089	0.086	0.084
	7	0.108	0.107	0.105	0.103	0.101	0.099	0.097	0.095	0.093	0.092
	8	0.103	0.102	0.102	0.101	0.100	0.100	0.099	0.098	0.098	0.097

The longer the ‘gradient’ the more spurious structure appears in higher dimensions

ARCHERY FOR EVERYONE

The literature on this effect is quite diverse, mostly in

→ ecology (e.g. Legendre & Legendre, 2003; Podani & Miklós, 2002)

→ applied statistics (e.g. Baccini et al., 1994; Diaconis et al., 2008)

and surprisingly inconclusive

Implications

→ Arches seem inevitable for text scaling problems

Why? Because Zipf!

→ 0 or 1 counts dominate document term matrices, so lack of overlap is almost guaranteed

COMPARISON

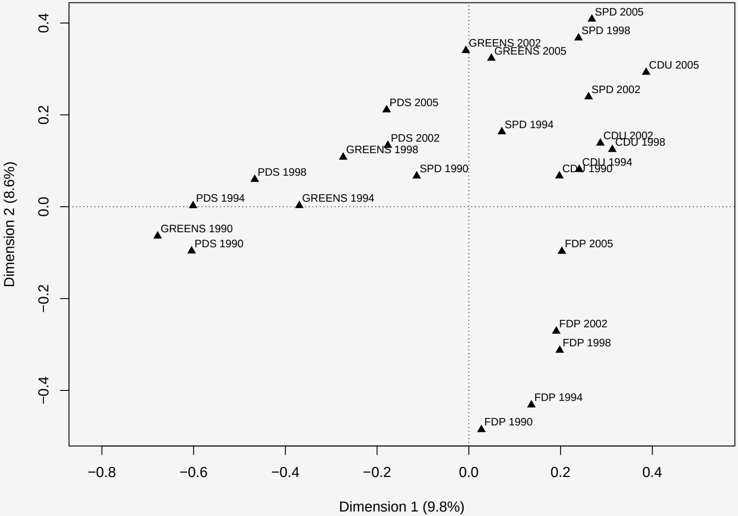
Two models of German party platforms 1990-2013

→ in policy mentions (from Marpor)

→ in word counts (from Slapin and Proksch)

	1	2	3	4	5	6	7	8	9	10
topics	0.311	0.137	0.092	0.088	0.058	0.048	0.039	0.033	0.031	0.023
words	0.098	0.086	0.056	0.055	0.051	0.045	0.043	0.041	0.039	0.039

COMPARISON


SPD 2005
SPD 1998
GREENS 2002
GREENS 2005
CDU 2005
SPD 2002
PDS 2005
SPD 1994
CDU 2002
CDU 1998
PDS 2002
GREENS 1998
CDU 1994
CDU 1990
SPD 1990
PDS 1998
PDS 1994
GREENS 1994
GREENS 1990
PDS 1990
FDP 2005
FDP 2002
FDP 1998
FDP 1994
FDP 1990
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
Dimension 1 (9.8%)
−0.4
−0.2
0.0
0.2
0.4
Dimension 2 (8.6%)

WHAT TO DO?

Some approaches:

→ Avoidance: scale aggregates, e.g. category or topic counts (e.g. Baerg & Lowe, 2020)
→ Correction: by 'detrending' CA (Hill, 1974; Wartenberg et al., 1987)
→ Constraint: by adding information about document positions (Laver et al., 2003; Palmer, 1993; ter Braak, 1986)
→ Agenda control: reduce ambition, pursue substantively meaningful scales one by one, e.g. Wordscores

Denial, anger, bargaining, depression

…acceptance?

HOW TO GET THE SCALE YOU WANT

From measurement theory

→ The meaning of a scales is identified by its *items*[1] and their loadings

The unique problem for text scaling is the sheer quantity of possible items

→ Often ~ 5 Likert items per construct

→ Marpor policy topic set: ~140 (we'll usually think about 56 main, or even 26 RILE)

→ German manifesto vocabulary: ~9000

Text also affords some novel *indirect* methods for choosing and loading items

→ Mostly to do with creative sample selection

[1]and not the damn logit transformation (you know who you are)

A TYPOLOGY OF ITEM STRATEGIES

Choosing (selection vs. using everything) and *loading* (selection vs. estimation) items

→ These are not quite separable, as loading an item with 0 is the same as not selecting it

Let's start with direct methods…

→ Use all items, select the loadings (e.g. Laver & Garry, 2000)

→ Select items, select loadings (e.g. RILE, Budge et al., 1987; Lowe et al., 2011) but also topic model-based scales and polar dictionary sentiment measures (Rauh, 2018)

→ Select items, estimate loadings (e.g. Lowe, 2016)

INDIRECT METHODS

The life-changing magic of *changing the unit of analysis*

- → Select only substantively relevant *document sections* (Proksch & Slapin, 2009; Slapin & Proksch, 2008)
- → Indirect control over word appearance and frequency
- → Indirect control over dimensionality

Seldom explored alternative: scale within estimated topic segmentation

AN ASIDE

The life-changing magic of *changing the unit of analysis*

- → Construct a 'word-document' for word j: local lexical contexts of j, summed
- → Fit scaling model with hundreds of dimensions
- → Set document/word positions to be $(\theta_i + \beta_j)/2$
- → Note: the matrix with 'word-document' vectors as rows is *symmetric* (only implementation regularization introduces differences)
- → Each multidimensional position is a *word embedding*

(Pennington et al., 2014)

INDIRECT METHODS

Add information about document *positions*

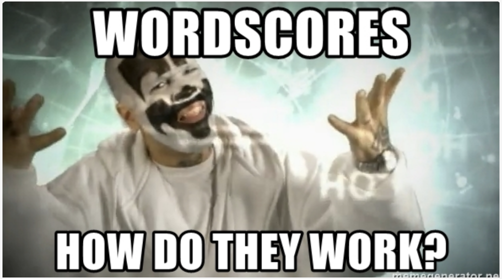
- → Wordscores (Perry & Benoit, 2017, and related methods, e.g.)
- → Row association model
- → Via document covariates

DOCUMENT POSITIONS AS MAGNETS

Wordscores

- → Assign scores to *two* reference documents, for simplicity -1 and 1
- → Reference documents select the item *set*
- → Compute item *loadings* 'word scores' for the vocabulary in those documents as the scores that, when averaged within each document, best recover their positions
- → Assign scores to other documents using those scores (out-of-sample)

A special case of the row association model (Goodman, 1979)



(I think we do not know how they work)

DOCUMENT POSITIONS AS MAGNETS

Wordscores

- → Assign scores to *two* reference documents, for simplicity -1 and 1
- → Reference documents select the item set
- → Compute loadings 'word scores' for the vocabulary in those documents as the scores that, when averaged within each document, best recover their positions
- → Assign scores to other documents using those scores (out-of-sample)

A special case of the row association model (Goodman, 1979)

Intuition (such as it is)

Reference document differences are

- → all due to variation on the the dimension of interest, for item selection and loading
- → reflected by differential word rates
- → which are represented by the loadings 'word scores'

Open questions

- → More than two reference scores?
- → Assumptions about other dimensions?
- → What is the role of item *selection*

ADDING INFORMATION INDIRECTLY

Document covariates

→ Canonical correspondence analysis (ter Braak, 1986)

→ A ‘structural’ scaling model

Two functions:

→ Force document positions to be in the *column space* of some document covariates

→ Force document positions to be in the *null space* of some document covariates

Promising, I think!

WHAT WAS ALL THAT ABOUT?

We can fit multidimensional text scaling models easily and cheaply using correspondence analysis

Results need careful interpretation, like all exploratory measurement tools

Orthogonal dimensions and ideal point structure requirements combine to generate recognizable artifacts: ‘the arch’

Solutions are as yet unclear, but may include

→ Avoiding the ‘long gradients’ that exacerbate it

→ Scaling substantive dimensions separately, by careful item selection

Existing item selection and weighting methods used in text scaling often work but we’re not quite sure how

Grounding scales using document covariates, e.g. using canonical correspondence analysis, seems like a natural solution

THANK YOU FOR YOUR ATTENTION

Llooking forward to your questions

REFERENCES

Baccini, A., Caussinus, H., & De Falguerolles, A. (1994). “Diabolic horseshoes.” *9th International Workshop on Statistical Modelling, Exeter*, 11–15.

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.

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.

Clinton, J., Jackman, S., & Rivers, D. (2004). “The statistical analysis of roll call data.” *American Political Science Review*, *98*(02), 1–16.

Diaconis, P., Goel, S., & Holmes, S. (2008). “Horseshoes in multidimensional scaling and local kernel methods.” *The Annals of Applied Statistics*, *2*(3), 777–807.

Elff, M. (2013). “A dynamic state-space model of coded political texts.” *Political Analysis*, *21*(2), 217–232.

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

Goodman, L. A. (1981). "Association models and canonical correlation in the analysis of cross-classifications having ordered categories." *Journal of the American Statistical Association*, *76*(374), 320–334.

Greenacre, M. (2007). "Correspondence analysis in practice." Chapman and Hall/CRC.

Hill, M. O. (1974). "Correspondence analysis: A neglected multivariate method." *Applied Statistics*, *23*(3), 340–354.

König, T., Marbach, M., & Osnabrügge, M. (2017). "Left/Right or U? estimating the Dimensionality of National Party Competition in Europe." *The Journal of Politics*, *79*(3), 1101–1105.

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.

Laver, M., & Garry, J. (2000). "Estimating policy positions from political texts." *American Journal of Political Science*, *44*(3), 619–634.

Legendre, P., & Legendre, L. (2003). "Numerical ecology (2nd English Ed.). Elsevier.

Lowe, W. (2016). *Scaling things we can count (formerly 'Putting it all on the line')*.

REFERENCES

Lowe, W., Benoit, K. R., Mikhaylov, S., & Laver, M. (2011). "Scaling policy preferences from coded political texts." *Legislative Studies Quarterly*, *36*(1), 123–155.

Palmer, M. W. (1993). "Putting things in even better order: The advantages of Canonical Correspondence Analysis." *Ecology*, *74*(8), 2215–2230.

Pennington, J., Socher, R., & Manning, C. (2014). "Glove: Global vectors for word representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Perry, P. O., & Benoit, K. (2017, October 24). *Scaling Text with the Class Affinity Model*. Retrieved October 19, 2021, from http://arxiv.org/abs/1710.08963

Podani, J., & Miklós, I. (2002). "Resemblance coefficients and the horseshoe effect in principal coordinates analysis." *Ecology*, *83*(12), 3331–3343.

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.

Rauh, C. (2018). "Validating a sentiment dictionary for German political language – a workbench note." *Journal of Information Technology & Politics*, *15*(4), 319–343.

REFERENCES

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.

ter Braak, C. J. F. (1986). “Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis.” *Ecology*, *67*(5), 1167–1179.

Turner, H., & Firth, D. (2008, October 16). *Generalized nonlinear models in R: An overview of the gnm package*. University of Warwick.

Wartenberg, D., Ferson, S., & Rohlf, F. J. (1987). “Putting things in order: A critique of detrended correspondence analysis.” *The American Naturalist*, *129*(3), 434–448.