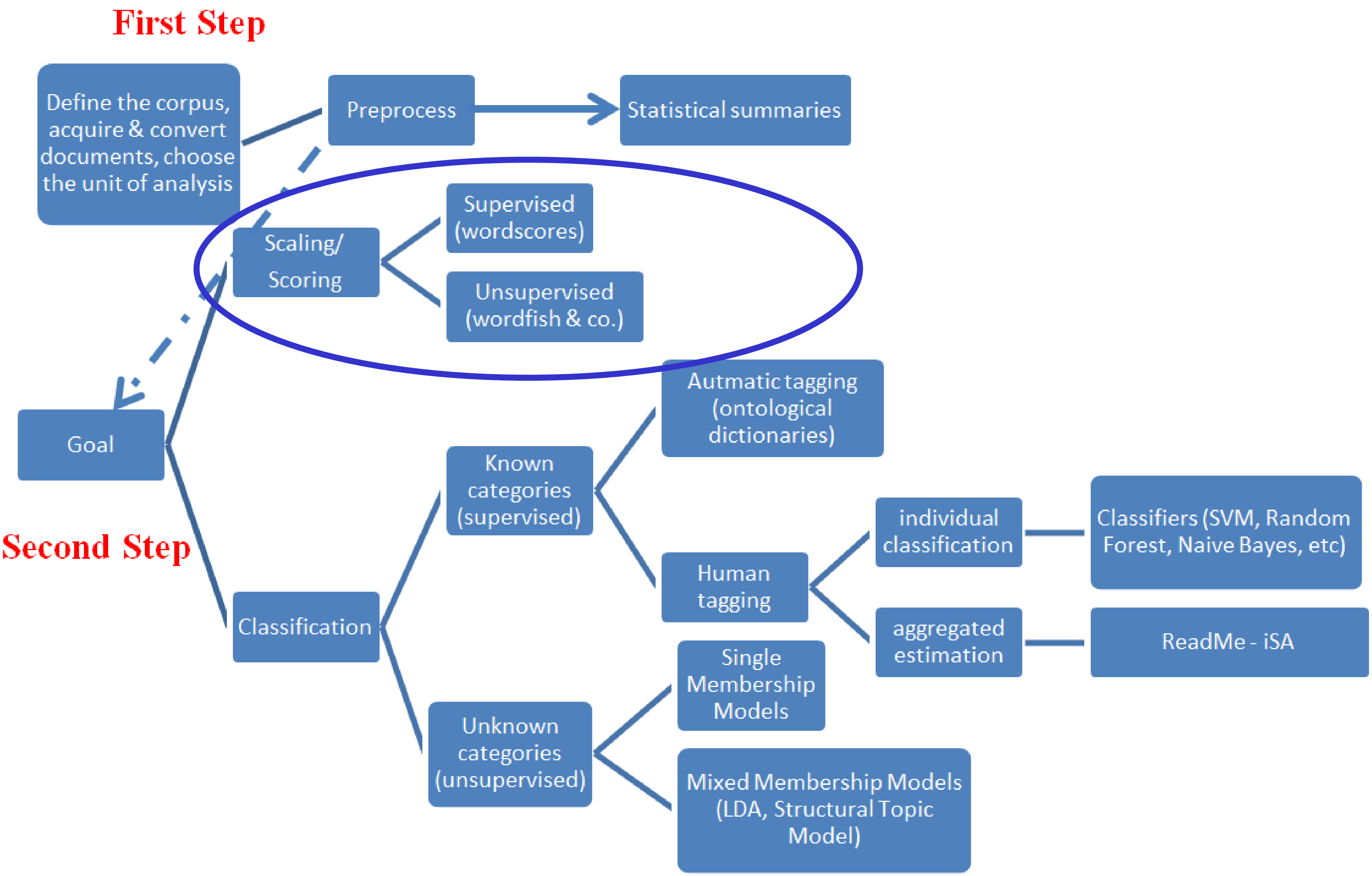
Applied Scaling & Classification Techniques in Political Science

Lecture 4 (part 1)
Further scaling algorithms

Our Course Map



First Step

Define the corpus, acquire & convert documents, choose the unit of analysis

Preprocess

Statistical summaries

Scaling/ Scoring

Supervised (wordscores)

Unsupervised (wordfish & co.)

Goal

Second Step

Classification

Known categories (supervised)

Autmatic tagging (ontological dictionaries)

Human tagging

individual classification

Classifiers (SVM, Random Forest, Naive Bayes, etc)

aggregated estimation

ReadMe - iSA

Unknown categories (unsupervised)

Single Membership Models

Mixed Membership Models (LDA, Structural Topic Model)

Other scaling algorithms

Wordshoal (an extension of Wordfish)

Suppose you want to scale legislative speeches to infer the position of legislators via an unsupervised scaling method

Political scientists have followed one of the two approaches when attempting to **recover preferences from legislative speeches**

Other scaling algorithms

Wordshoal (an extension of Wordfish)

One approach has been to confine the analysis to speeches on a **single legislative act**, such as a motion of confidence or during the general policy speech of the PM (the approach we saw in the Japanese analysis of legislative speeches via Wordfish)

While this approach (by assumption) holds **topical variation constant**, the resulting estimates are confined to the set of legislators who spoke and **the topic** on which they spoke

Other scaling algorithms

Wordshoal (an extension of Wordfish)

The opposite approach has been to **combine many speeches over many legislative acts** into a single document for each legislator

For example, scale speeches from the European Parliament by aggregating contributions across many topics **by national parties**

By **pooling speeches across many topics**, these authors have implicitly hoped that different parties would each discuss a similar mixture of topics, and therefore topical variation would cancel out

Other scaling algorithms

Wordshoal (an extension of Wordfish)

While this can work at the party level, topical mixes **vary enormously** at the level of individual speakers. How to deal with that?

How to deal with this challenge?

Other scaling algorithms

Wordshoal (an extension of Wordfish)

Wordshoal (a “shoal” is a group of fish, not traveling in the same direction)!

Wordshoal is based on two stages:

The first stage uses “Wordfish” to scale word use variation in **each debate separately**

In the second stage, it uses Bayesian factor analysis to construct a **common scale** from the debate specific positions estimated in the first stage

Other scaling algorithms

Wordshoal (an extension of Wordfish)

One of Worshoal’s key innovations is that it allows the meaning and **discriminatory power of a given word to vary from debate to debate**

For example, the word “debt” may be important to discriminate speakers in a debate on extending health care, while the same word may have little discriminatory power in a debate on the budget deficit, where it will be used heavily by most speakers

Other scaling algorithms

Wordshoal (an extension of Wordfish)

Because this approach does not rely on word use variation in any single debate to estimate positions on a latent dimension of disagreement, it gains **additional robustness** against other sources of variation in word usage

All we need to discover this latent dimension is for that dimension to have **general predictive power for word use variation** across the set of observed debates

Other scaling algorithms

Wordshoal (an extension of Wordfish)

To stick with our example, Wordshoal allows you to scaling the entire speech corpus of a legislature, producing valid legislator-specific scores on some underlying general dimension(s) that can be used to study, for example, legislative behavior, intra-party politics, and polarization

This of course is not true only for legislative speeches! Wordshoal is **attractive everytime** you want to analyze several different speeches/documents per-speaker/actor taken in very different contexts (over possible different topics)

Other scaling algorithms

Wordshoal (an extension of Wordfish)

For example, suppose we have party programs over a long time-period and many elections

As we saw in the German example, we have a **temporal challenge** here: the language, but ALSO the topics covered in the different elections, will tend to vary. As a result, documents' positions over time could be not reliable!

How to deal with this challenge? Either going with the strategy we already discussed in Lecture 2 (by reducing the number of features included in the DfM), or using Wordshoal on the entire DfM!!!

Other scaling algorithms

Wordshoal (an extension of Wordfish)

Lauderdale, Benjamin E., and Alexander Herzog (2016). Measuring Political Positions from Legislative Speech, *Political Analysis* (2016) 24:374–394

```
devtools::install_github("kbenoit/wordshoal")
```

Quanteda command: `textmodel_wordshoal`

Other scaling algorithms

Class affinity (an extension of Wordscores)

Class affinity is attractive every time you have a few examples of documents at the extremes of a hypothesized ideological or stylistic spectrum and you want to estimate the probability (the degree of similarity) of your set of documents/texts to belong to one out of two categories (0/1, Government/Opposition, etc.)

Perry, P.O. & Benoit, K.R. (2017). Scaling Text with the Class Affinity Model. arXiv:1710.08963 [stat.ML]

Quanteda command: `textmodel_affinity`

Other scaling algorithms

Class affinity (an extension of Wordscores)

The basic conceptual behind a class affinity model:

- ✓ Over the course of a speech, for example, a speaker orientation **switches back and forth** between Government mode and Opposition mode
- ✓ When she is in Government mode, she chooses words in the same manner as the government leadership
- ✓ Likewise, when she is Opposition mode, she chooses words in the same manner as the opposition leadership

Other scaling algorithms

Class affinity (an extension of Wordscores)

- ✓ We should therefore place the speaker on the spectrum between the two extremes of pro-government and pro-opposition according to what **proportion of time she spends in each mode**
- ✓ In the class affinity framework, this learning step requires not large volumes of training data, but rather texts that are clearly polar examples of each reference class (also **more than ones at each extreme**), to form benchmarks for estimating the other texts' affinities to these classes

Other scaling algorithms

Correspondence Analysis (old stuff, good stuff!)

As in principal component analysis, the idea in CA is to reduce the dimensionality of a data matrix and visualize it in a subspace of low-dimensionality

The data of interest in simple CA are usually a two-way contingency table for which ***relative*** (not absolute) values are of primary interest

Other scaling algorithms

Correspondence Analysis (old stuff, good stuff!)

The issue on “relative emphasis” makes (political) sense!

Budge (2001) lays out the basic assumptions of the saliency theory of party competition. According to his theory “policy differences between parties consist of contrasting emphases on different policy areas (thus, one party often mentions taxes, another benefits)”

That is, it is only the “**relative emphasis**” of one word over another that signals position

Other scaling algorithms

Correspondence Analysis (old stuff, good stuff!)

Lowe (2008; 2016) shows that CA provides an approximation to a Poisson ideal point model for text data (i.e., Wordfish!)

In most applications it does **not make much difference** which model is used; however, it has been found that Wordfish is **more robust** when a single document is **very different** than the others, which happens not infrequently in political documents

So why do we bother to CA after all if we have Wordfish?!?

Other scaling algorithms

Correspondence Analysis (old stuff, good stuff!)

The advantages of CA:

- ✓ You can use it as a diagnostic tool to understand if 1-dimensional world is a good approximation of the informatiotion included in your data, or not. Remember that via Wordfish you can get only 1 dimension
- ✓ You can easily estimate a 2-dimensional world

The limits of CA:

- ✓ no uncertainty estimation
- ✓ validating the latent space extracted it more tricky than Wordfish

Other scaling algorithms

What is a Correspondence Analysis?

Let’s give a non-(too)-technical explanation via R