Content Analysis Dictionaries 2

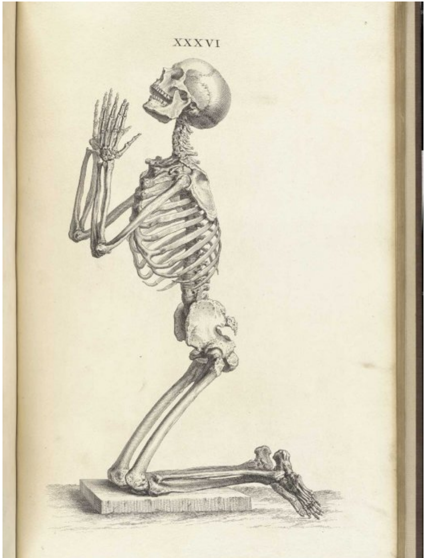
William Lowe

Hertie School

28th September 2020

Solutions: Some theological approaches

Solutions: Some theological approaches



How bad is it?

- → *Recall*: the proportion of words used that way that are in your dictionary
- → *Precision*: the proportion of words used the way your dictionary assumes they are used

Precision and recall

Every field reinvents this distinction:

→ precision and recall

→ specificity and sensitivity

→ users and producer's accuracy

→ type 1 and type 2 error

→ sins of omission and sins of commission

Precision

Keyword in context analyses (KWIC) allow you to scan all contexts of a word

→ How many of them *are* the sense or usage you want?

Let's take a look at `benefit*` as a 'pro government intervention in the economy' word

	pre	keyword	post
1	also keep all the other	benefits	that pensioners currently receive ,
2	regulation will have to have	benefits	exceeding costs , and regulations
3	and Controlled Immigration Britain has	benefited	from immigration . We all
4	positive contribution But if those	benefits	are to continue to flow
5	Northern Ireland brings	benefits	to all parts of our
6	their home , will also	benefit	first-time buyers . Empowering individuals
7	you help yourself ; you	benefit	and the country benefits .
8	you benefit and the country	benefits	. So now , I
9	result of our tax and	benefit	measures compared to 1997 .
10	result of personal tax and	benefit	measures introduced since 1997 ,
11	, the savings on unemployment	benefits	will go towards investing more
12	trebled the number on incapacity	benefits	. We will help 17
13	Work programme and reform Incapacity	Benefit	, with the main elements
14	main elements of the new	benefit	regime in place from 2008
15	stronger penalties . To the	benefit	of business and household consumers
16	effective directive to provide real	benefits	to consumers and new opportunities
17	better.We are examining the potential	benefits	of a parallel Expressway on
18	ways to lock in the	benefit	of new capacity . We
19	are determined to spread the	benefits	of enterprise to every community
20	to get ahead , to	benefit	from improving public services

Precision

Of the 20 instances, these are (arguably)

- → 6 used the way we expect from the topic
- → 3 used in the opposite sense: anti-government intervention in the economy
- → 11 used in ways that are neither

So… 0.3 correct, 0.15 mistaken, and 0.55 unrelated 'noise'

- → Perhaps not an amazing choice

There are two kinds of precision failures here with different consequences

- → Mistaking an *topic-unrelated* word for this topic (11 of these)
- → Mistaking a word used in the sense of a *different* topic for this one (3 of these)

The first mistake does not really harm precision, but the second does

Recall

Bad recall is a mixture of two problems

- Assigning words to a topic that are mostly used for a different one: $P(W \mid Z = k) < P(W \mid Z = j)$ but we assigned it to k anyway
- Failing to assign a topic-informative word to any topic: Dictionary says $P(W \mid Z = k) = 0$, but it's not. This is about *coverage*

Let's consider coverage first

Coverage

One possible checking procedure:

→ Take a random matched sample of words not in the dictionary but present in the corpus e.g. match each dictionary word to another of the same frequency

→ Examine their KWICs to see if they should have been assigned to a topic

If we were feeling even more energetic

→ Assign them their most likely topic manually

→ Compare this $\tilde{\theta}$ to the dictionary's own estimate θ

→ These should not be wildly different

Recall

The other kind of mistake is difficult because the natural procedure is

- → Assign *every word* (or at least every instance of a word that the dictionary knows about in a document) to a topic
- → For each topic, see what proportion of times the dictionary agrees it is in that topic

This also promises to be very tiring.

Using precision to estimate recall

However, two facts may help us:

- → We have a tireless computer available
- → Recall and precision relate $P(W \mid Z)$ and $P(Z \mid W)$ respectively
- → …and we know how

Let's call the *true* topic of a word Z as before, and the *dictionary's idea of the topic* of a word $\hat{Z}$ because it's kind of an estimate of that. So,

$$\text{Recall:} \quad \sum_k^K P(\hat{Z} = k \mid Z = k)$$

$$\text{Precision:} \quad \sum_k^K P(Z = k \mid \hat{Z} = k)$$

Using precision to estimate recall

If we have a sense of dictionary topic precision, could we use it to get a sense of dictionary topic recall?

Using precision to estimate recall

If we have a sense of dictionary topic precision, could we use it to get a sense of dictionary topic recall? Why yes, by borrowing methods from King and Lowe (2003).

Using precision to estimate recall

If we have a sense of dictionary topic precision, could we use it to get a sense of dictionary topic recall? Why yes, by borrowing methods from King and Lowe (2003).

According to the Rev. Bayes

$$\begin{aligned} P(\hat{Z} = k \mid Z = k) &= \frac{P(Z = k \mid \hat{Z} = k)P(\hat{Z} = k)}{\sum_j^K P(Z = j \mid \hat{Z} = j)P(\hat{Z} = j)} \\ &= P(Z = k \mid \hat{Z} = k)\frac{P(\hat{Z} = k)}{P(Z = k)} \qquad \text{(recall is reweighted precision)} \\ &\propto P(Z = k \mid \hat{Z} = k)P(\hat{Z} = k) \end{aligned}$$

Conveniently

- → we don't need the denominator because it only ensure the recall measures add to one
- → We can get $P(\hat{Z} = k)$ by running the dictionary over the entire corpus

Perspective

Precision and recall are useful, but we are most interested in the *consequences* of being wrong

Previously: consequences for a left-right ideology measure from Laver and Garry (2000)

Let's look at a related conflict-cooperation score assigned by experts to dictionary 'topics' (King & Lowe, 2003)

A machine coding system 'read' Agence France Press leads on the Bosnian wars and generated event data

- → Who did what to whom when

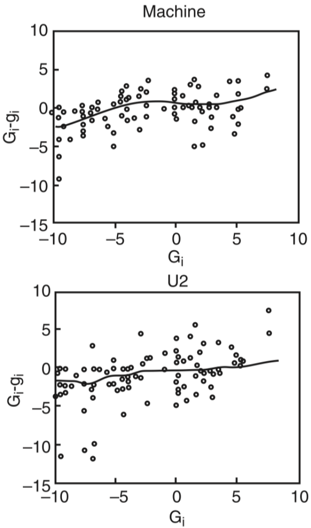
We were interested in evaluating the 'what'

- → Topics are event types in a large ontology of international events
- → topic *scores* represent how conflictual or cooperative that event is
- → Some topic categorization mistakes will matter more than others

Event topics

Goldstein	*IDEA*	*Definition*	*Goldstein*	*IDEA*	*Definition*
0.1	091	ask for information	−7.6	1826	military border fortification
0.1	024	optimistic comment	−7.6	1825	military mobilization
0	99	sports contest	−7.6	1824	military troops display
0	98	A and E performance	−7.6	1823	military naval display
0	97	accident	−7.6	1821	military alert
0	96	natural disaster	−7.6	182	military demonstration
0	95	human death	−8.3	224	riot or political turmoil
0	94	human illness	−8.7	221	bombings
0	72	animal death	−9.2	2236	military seizure
0	27	economic status	−9.2	2123	abduction
0	26	adjust	−9.2	211	seize possession
0	25	vote	−9.6	2228	assassination
0	24	adjudicate	−9.6	2227	guerrilla assault
0	2321	government default on payments	−9.6	2226	paramilitary assault
0	2312	private transactions	−9.6	2225	torture
0	2311	government transactions	−9.6	2224	sexual assault
0	231	transactions	−9.6	2223	bodily punishment
0	23	economic activity	−9.6	2222	shooting
−0.1	094	ask for protection	−9.6	2221	beatings
−0.1	022	pessimistic comment	−9.6	222	physical assault
−0.1	021	decline comment	−9.6	22	force
−0.1	02	comment	−10	2237	biological weapons use
−0.9	141	deny responsibility	−10	2235	assault
−1	14	deny	−10	2234	military occupation
−1.1	0631	grant asylum	−10	2233	coups and mutinies
−2.2	192	reduce routine activity	−10	2232	military raid
−2.2	121	criticize or blame	−10	223	military engagements
−2.4	132	formally complain			

Topic scores



Topic scores

Not bad (or at least not bad relative to undergraduate coders)

All the evaluation principles we've seen here apply to

- → document classifiers
- → content analysis dictionaries
- → human coders

Nothing is more practical than a good theory…

Topic scores

Not bad (or at least not bad relative to undergraduate coders)

All the evaluation principles we've seen here apply to

- → document classifiers
- → content analysis dictionaries
- → human coders

Nothing is more practical than a good theory…

Next week we'll see how we learn dictionaries rather than write them

- → but all the same considerations will apply

References

King, G. & Lowe, W. (2003). 'An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design'. *International Organization*, *57*(3), 617–642.

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

Mikhaylov, S., Laver, M. & Benoit, K. R. (2011). 'Coder reliability and misclassification in the human coding of party manifestos'. *Political Analysis*, *20*(1), 78–91.