8 Cases:

* True Reliable
  + Case 1: BERT ICA Agree on Reliable
  + Case 2: BERT ICA agree on Unreliable
  + Case 3: Bert thinks Reliable and ICA thinks Unreliable
  + Case 4: BERT thinks Unreliable and ICA thinks Reliable
* True Unreliable
  + Case 5: BERT ICA Agree on Unreliable
  + Case 6: BERT ICA agree on Reliable
  + Case 7: Bert thinks Unreliable and ICA thinks Reliable
  + Case 8: BERT thinks Reliable and ICA thinks Unreliable

Case 1:

Tweet 462: WORLD: Air Canada cancels select flights to China as coronavirus spreads

True class reliable:

BERT Reliable; ICA Reliable

Observations: All values are very small for both BERT and ICA (as compared to Case 4). Question: are both systems unsure (low values?)

ICA tends to think the words tend toward unreliable (except for spreads and cancels, which are on the reliable side). BERT put more weight on unreliable (but less than ICA if the two measures are comparable)… Despite the tendency toward unreliable, they both classify this tweet as reliable… Again, maybe because of the small values which hover around 0.

Case 2:

Is the coronavirus a threat to the Canadian economy? We chat with Ian Lee of @Carleton\_U Sprott School of Business.

True class reliable:

BERT Unreliable; ICA Unreliable

Observations: For BERT: values do hover around 0 and Carleton\_U’s reliability should have compensated for Sprott’s unreliabiliety. But We is very high. ICA has several values at 0.05. Doesn’t hover around 0 enough? ICA is tricked by all the unkknowns (proper names). That makes sense. It doesn’t make sense for BERT not to be tricked by Carleton\_U or to think that We is dangerous… They both fail, but I feel ICA has an excuse.

Case 3:

Metrolinx installing hand sanitizers on GO Transit buses and inside stations to stop spread of coronavirus #COVID19

True class reliable:

BERT Reliable; ICA Unreliable

Observations: Both systems tend toward unreliable, but, for BERT, values hover around 0 and the reliability of COVID19 compensates for the unreliability of, say, Metrilinx, so it played it safe? ICA wqs similar, but with values greater than BERT (but the compensation of COVID19 for stations is here). HOWEVER, it mistakenly gave a lot of unreliable value to stop word (which I include here because it is telling): “inside”. BERT also thinks it tends toward unreliable, but not as much.

Case 4:

Tweet 458: Fear and misinformation can lead to anxiety. Get the facts about #coronavirus. Don't be afraid to #BellLetsTalk

True class reliable:

BERT Unreliable; ICA Reliable

Graph:

Conclusion: ICA “understood” the context of the text (when afraid 🡺 Unknown (BellLetsTalk)). It gave more importance to what it knew (Fear and Anxiety are related and together they form a “reliable force” even if afraid is related and is a more “unreliable” term

BERT “interpreted the unknown (BellLetsTalk) as a threat, compounded it with misinformation and coronavirus along with Fear and afraid as unreliable.

Case 5:

True class unreliable:

BERT Unreliable; ICA Unreliable

My friend was at O’Hare today and heard people laughing at the idea of washing hands to prevent coronavirus.

Graph:

Observations: Unreliable trend by both, but ICA has three instances that are reliable for words that, indeed, should be reliable. In my mind, because I don’t believe that the tweet is truly unreliable (I contest the gold standard: the human rules are problematic too!), I am encouraged by the fact that ICA is marking some words as reliable… BERT, on the other hand, finds all the words unreliable. (Of course, if I find mislabelings in the testing gold standards, they are also present in the training one (or is this the result of cross-validation; i.e., we see all the data in both training and testing?)… In that case, it is not “wrong” of the system to do what it does.

Case 6:

Although there is now a presumptive confirmed case of Wuhan novel coronavirus, I want you to know that Ontario is prepared

True class unreliable:

BERT Reliable; ICA Reliable

Graph:

Discussion: Both systems picked Wuhan and presumptive as unreliable which is encouraging… However, they both classified as reliable. Not clear why BERT did: it seemed to trend toward unreliable. ICA seemed confused by this post: it seems to break even.

Case 7:

CNN Has Been Deliberately Trying To Incite Panic With #PresidentTrump Has No #CoronaVirus Plan Talking Points Watch

True class unreliable:

BERT Unreliable; ICA Reliable

Graph:

Observations:

ICA’s trend is unreliable, but is not strongly enough committed to it with many values around 0. Mistakes it makes seem to be that Panic is on the reliable side and Watch on the unreliable one (though are the words’ font respected? If so, the fact that watch is capitalized could lead it toward unreliable). CNN is rightly associated with reliability.

BERT is trending fully toward unreliable, though sometimes for the wrong reasons?: (PresidentTrump (unknown token), Has, Talking, Coronavirus, Trying CNN, Been).

Case 8:

Chinese scientists says COVID-19/coronavirus could have originated from a Gov Lab

True class unreliable:

BERT Reliable; ICA Unreliable

Graph:

Discussion: BERT hovers around 0 even if the trend is unreliable 🡪 reliable; ICA has strong “feelings” toward unreliable (Chinese, coronavirus, scientists, Lab and originated are strong unreliable words for ICA. Lab and Coronavirus stand out) 🡪 unreliable.

Note: this is a subtle tweet: it seems respectable, but it is its seeming respectability that makes it unreliable. Did ICA really catch that? How? Was it trained on a lot of data that was unreliable and talked about Chinese scientists, Labs and origination? I assume so. Why did BERT not pick up on that from the training data? Perhaps because ICA “thinks” in topics, whereas for BERT, words are just words? TThis is an interesting case… Perhaps it would be worth following up on it to see what happened (by checking the training data, but that’s a bit of a pain.