A benchmark dataset for detecting frames in multi-topical news content: Online Appendix

A benchmark dataset for detecting frames in multi-topical news content: Online Appendix

Comparing confidence level of correct and incorrect expert coding

We modeled the correctness of expert coding ("F1" is equal to the ground truth) and confidence level ("F2"), while adjusting for individual differences between the two experts using Bayesian multilevel logistic regression analysis. The following is the robust conditional effect plot. There is no evidence to suggest that there is a trend. Therefore, experts can either confidently give correct and incorrect coding.

Regression coefficients

Table 1

 # Fixed Effects

Parameter	Median	89% CI	pd	Rhat	ESS
(Intercept)	0.47	(0.38, 0.56)	99.92%	1.005	1091.00
method_typeSemisupervised	-0.11	(-0.22, 5.98e-03)	93.88%	1.005	1020.00
method_typeAutomatic	-0.15	(-0.26, -0.05)	98.02%	1.005	1037.00

Table 2

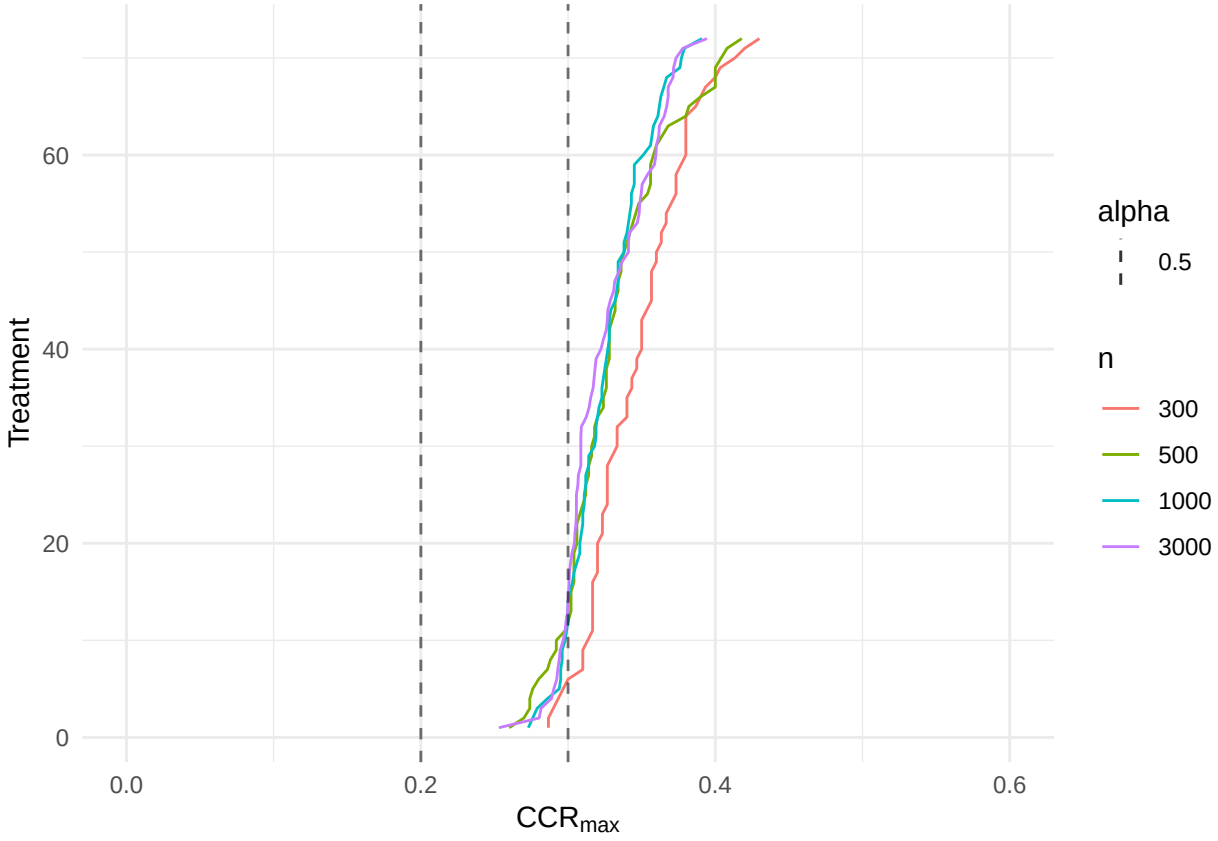
 # Sigma

Parameter	Median	89% CI	pd	Rhat	ESS
sigma	0.05	(0.05, 0.05)	100%	0.999	3924.00

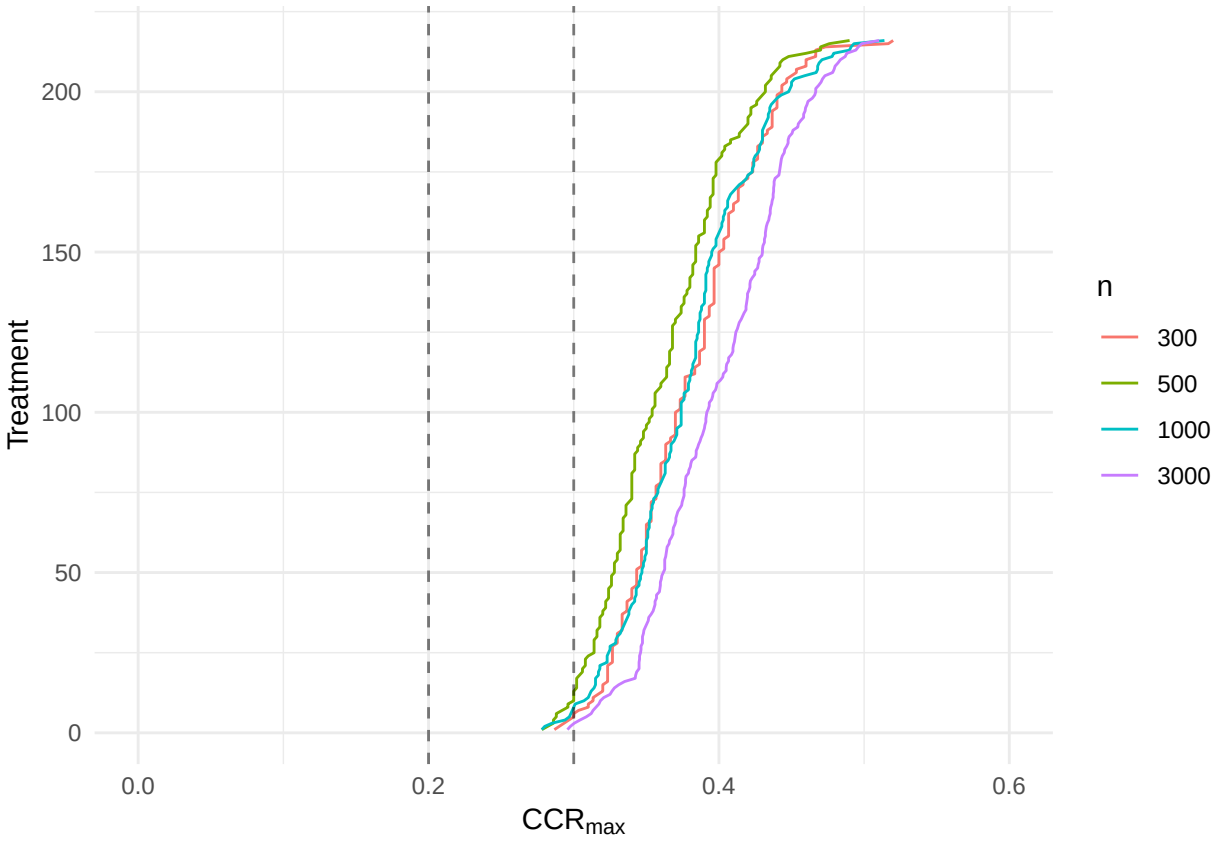
$R^2 = 0.1840042$

Simulation of increasing sample size

LDA.



Seeded-LDA.


200
150
100
50
0
Treatment
0.0
0.2
0.4
0.6
CCR$_{max}$
n
300
500
1000
3000

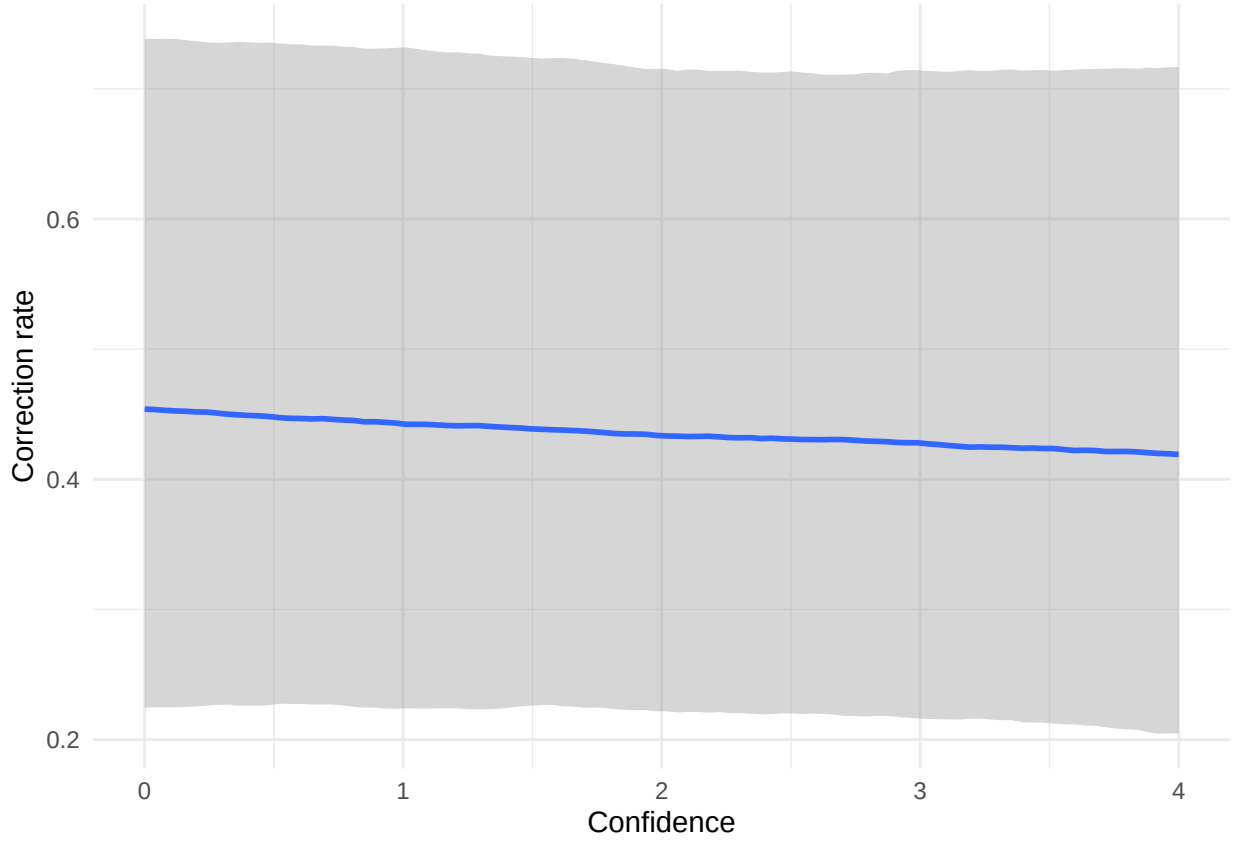


Figure 1. Robust conditional effects from the Bayesian model on the relationship between correction rate of expert coding and confidence at 89% credibility