An Interpretable Joint Graphical Model for Fact-Checking from Crowds

An T. Nguyen[1] Aditya Kharosekar[1] Matthew Lease[1]
Byron C. Wallace[2]

[1]University of Texas at Austin

[2] Northeastern University

Problems

Given a claim:

> *Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.*

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Predict headline stance: For Against Observing

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Predict headline stance:	For	Against	Observing
Predict claim veracity:	False	True	Unknown

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Predict headline stance:	For	Against	Observing
Predict claim veracity:	False	True	Unknown

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Predict headline stance:	For	Against	Observing
Predict claim veracity:	False	True	Unknown

Our motivation:

- Make sense of general claims incl. scientific, historical, ...

Problems

Given a claim:

Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language.

and relevant article headlines:

No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.

source: gizmodo.com

Predict headline stance:	For	Against	Observing
Predict claim veracity:	False	True	Unknown

Our motivation:

- Make sense of general claims incl. scientific, historical, ...
- Not just "fake news".

Solutions

Previous work:

Solutions

Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).

Solutions

Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).
- Predict veracity from stance+source features (Popat et al. 2017)

Solutions

Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).
- Predict veracity from stance+source features (Popat et al. 2017)

We proposed:

- Crowdsource stance labels.

Solutions

Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).
- Predict veracity from stance+source features (Popat et al. 2017)

We proposed:

- Crowdsource stance labels.
  - Hybrid human AI
  - Available near real-time

Solutions

Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).
- Predict veracity from stance+source features (Popat et al. 2017)

We proposed:

- Crowdsource stance labels.
  - Hybrid human AI
  - Available near real-time
- Joint graphical model of stance, veracity, annotators.

Solutions

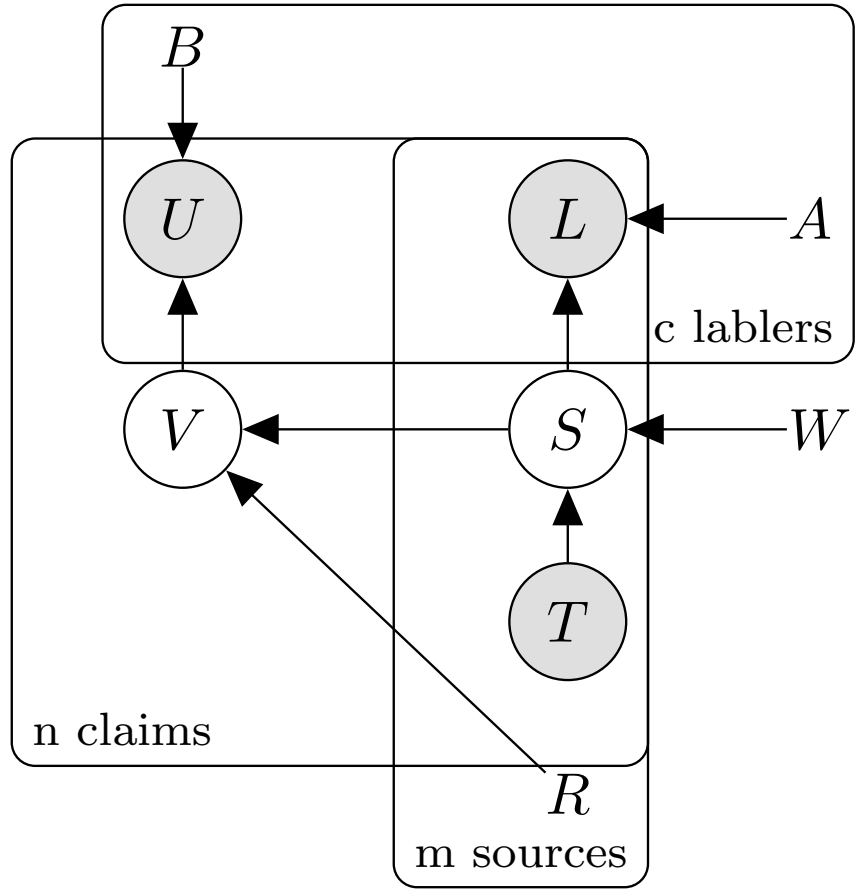
Previous work:

- Predict stance from text features (Ferreira& Vlachos 2016).
- Predict veracity from stance+source features (Popat et al. 2017)

We proposed:

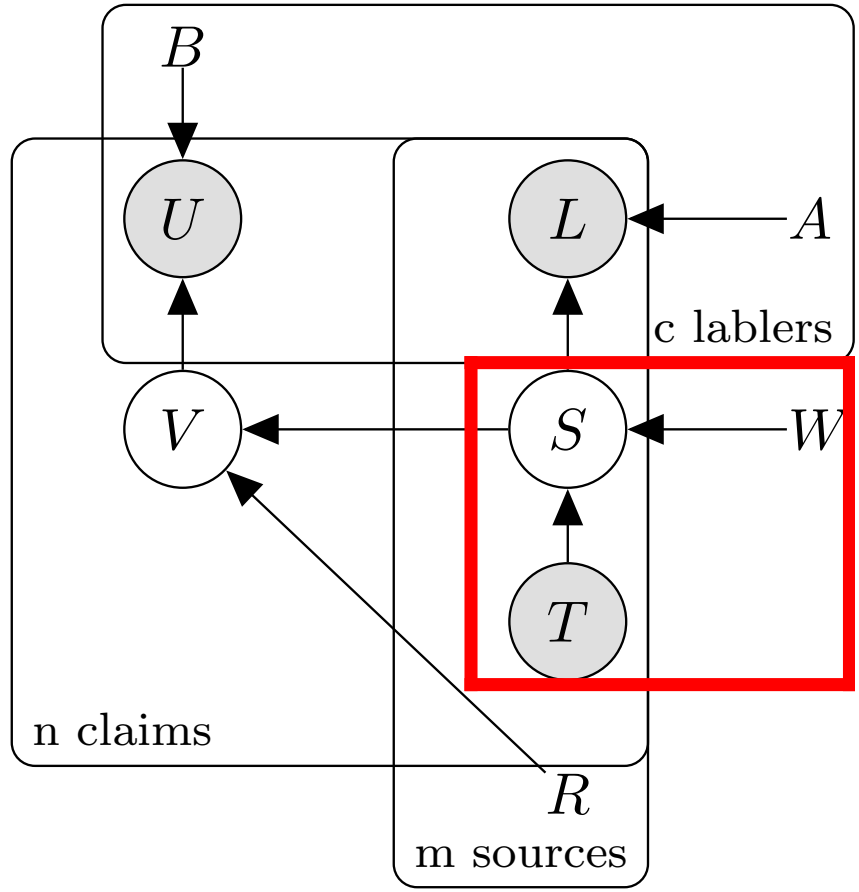
- Crowdsource stance labels.
  - Hybrid human AI
  - Available near real-time
- Joint graphical model of stance, veracity, annotators.
  - Interaction between variables
  - Interpretable

Model



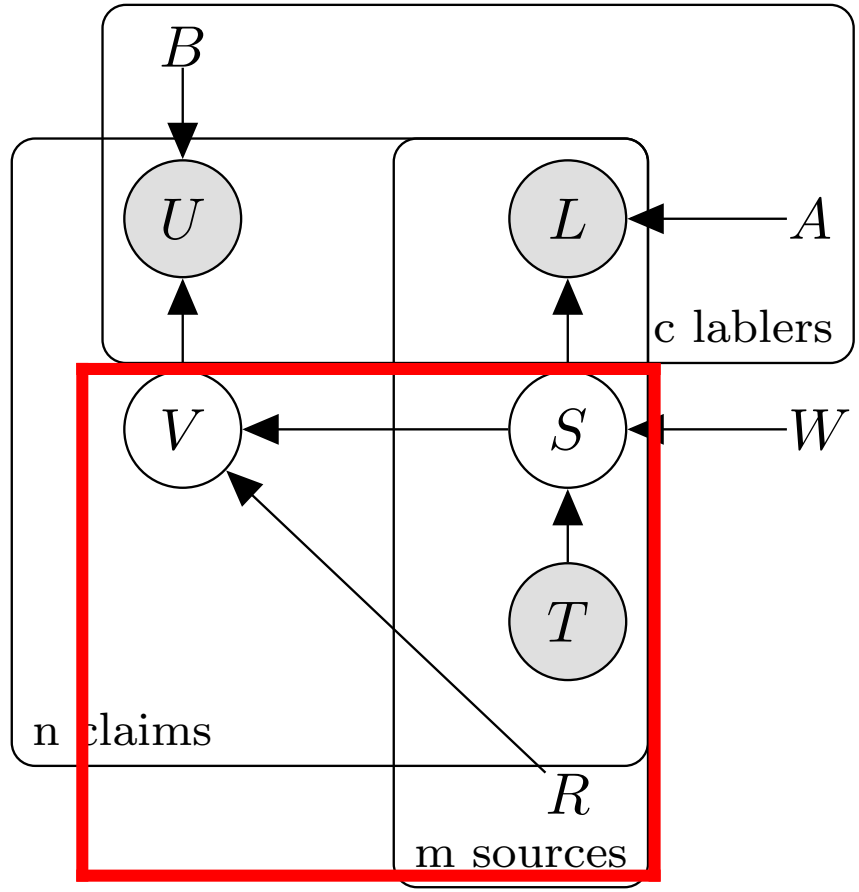
Model

1. Predict Stance S
   - Text features T



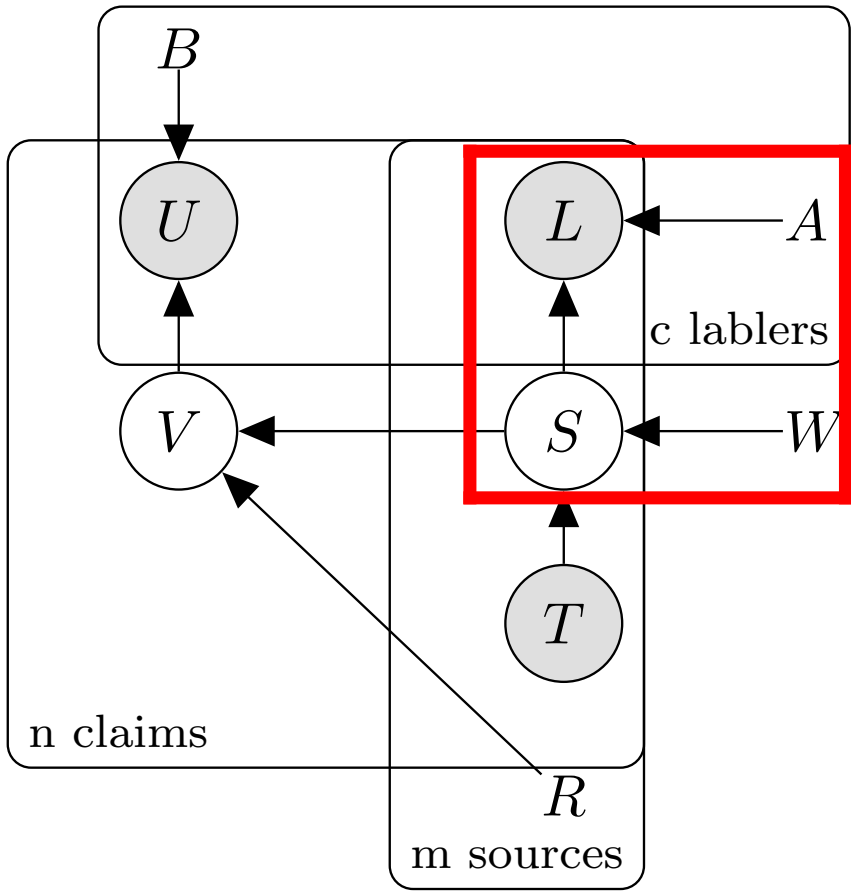
Model

1. Predict Stance S
   - Text features T
2. Predict Veracity V
   - Stance S
   - Reputation R



Model

1. Predict Stance S
   - Text features T
2. Predict Veracity V
   - Stance S
   - Reputation R
3. Stance Label L
   - True stance S
   - Annotator competence A



Inference & Learning

Inference:

- Gibbs sampling: accurate but slow.

Inference & Learning

Inference:

- Gibbs sampling: accurate but slow.
- Variational inference: fast but biased.

Inference & Learning

Inference:

- Gibbs sampling: accurate but slow.
- Variational inference: fast but biased.

Learning: Expectation Maximization.

Inference & Learning

Inference:

- Gibbs sampling: accurate but slow.
- Variational inference: fast but biased.

Learning: Expectation Maximization.

Details in the paper.

Evaluation

Data: Emergent (Ferreira and Vlachos 2016)

- 300 claims.
- 2595 articles with stance labels.

Evaluation

Data: Emergent (Ferreira and Vlachos 2016)

- 300 claims.
- 2595 articles with stance labels.
- We collected: crowd stance labels by Mechanical Turk.

Evaluation

Data: Emergent (Ferreira and Vlachos 2016)

- 300 claims.
- 2595 articles with stance labels.
- We collected: crowd stance labels by Mechanical Turk.

Baseline: **Separated** models for stance, veracity & crowd labels.

Evaluation

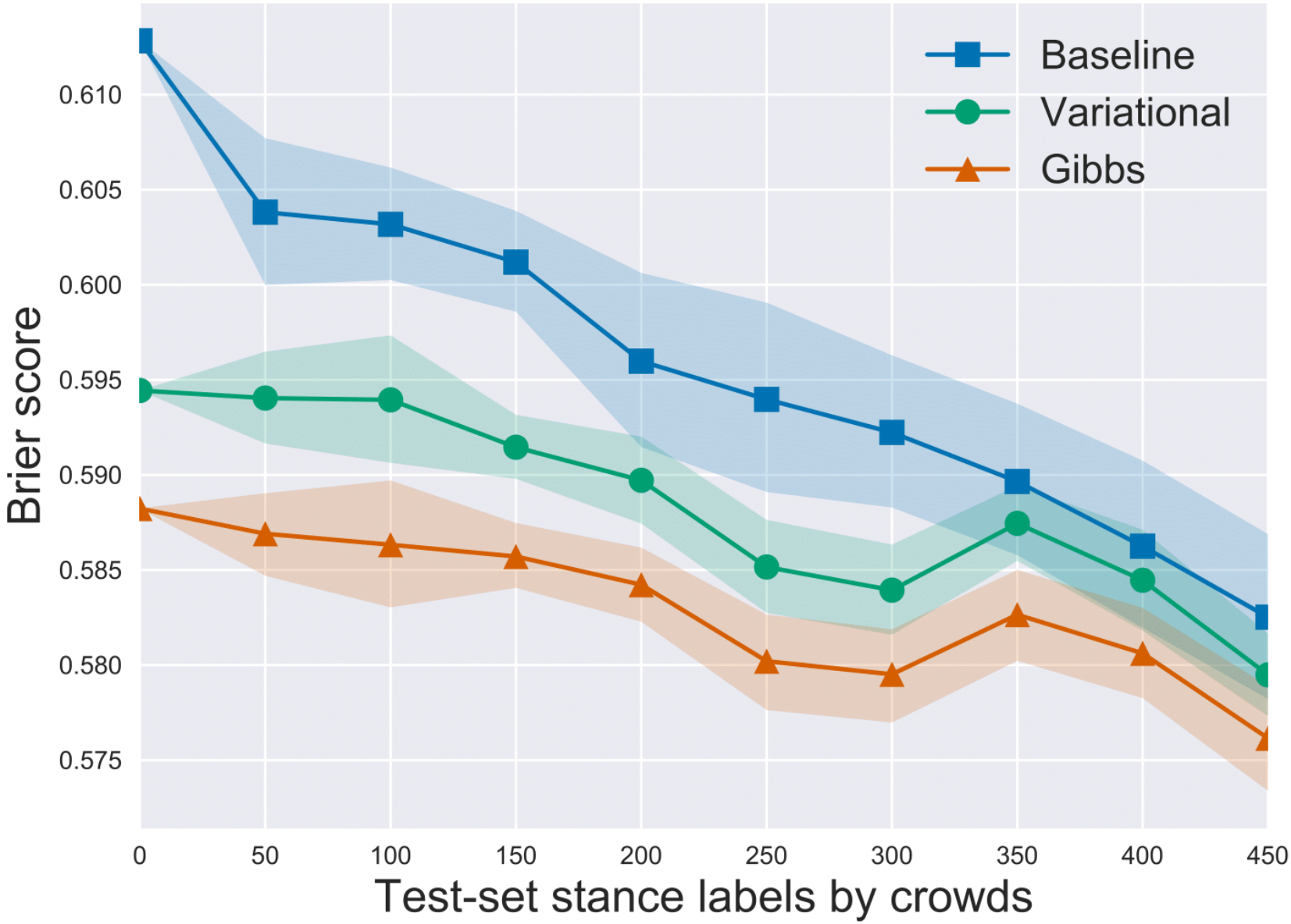
Data: Emergent (Ferreira and Vlachos 2016)

- 300 claims.
- 2595 articles with stance labels.
- We collected: crowd stance labels by Mechanical Turk.

Baseline: **Separated** models for stance, veracity & crowd labels.

Metric: Brier score, measures accuracy and prob. calibration.

Results


Baseline
Variational
Gibbs
Brier score
0.610
0.605
0.600
0.595
0.590
0.585
0.580
0.575
0
50
100
150
200
250
300
350
400
450
Test-set stance labels by crowds

User study

Interface: users enter claims, see predictions.

User study

Interface: users enter claims, see predictions.

A/B testing

User study

Interface: users enter claims, see predictions.

A/B testing

- A: see only veracity predictions

User study

Interface: users enter claims, see predictions.

A/B testing

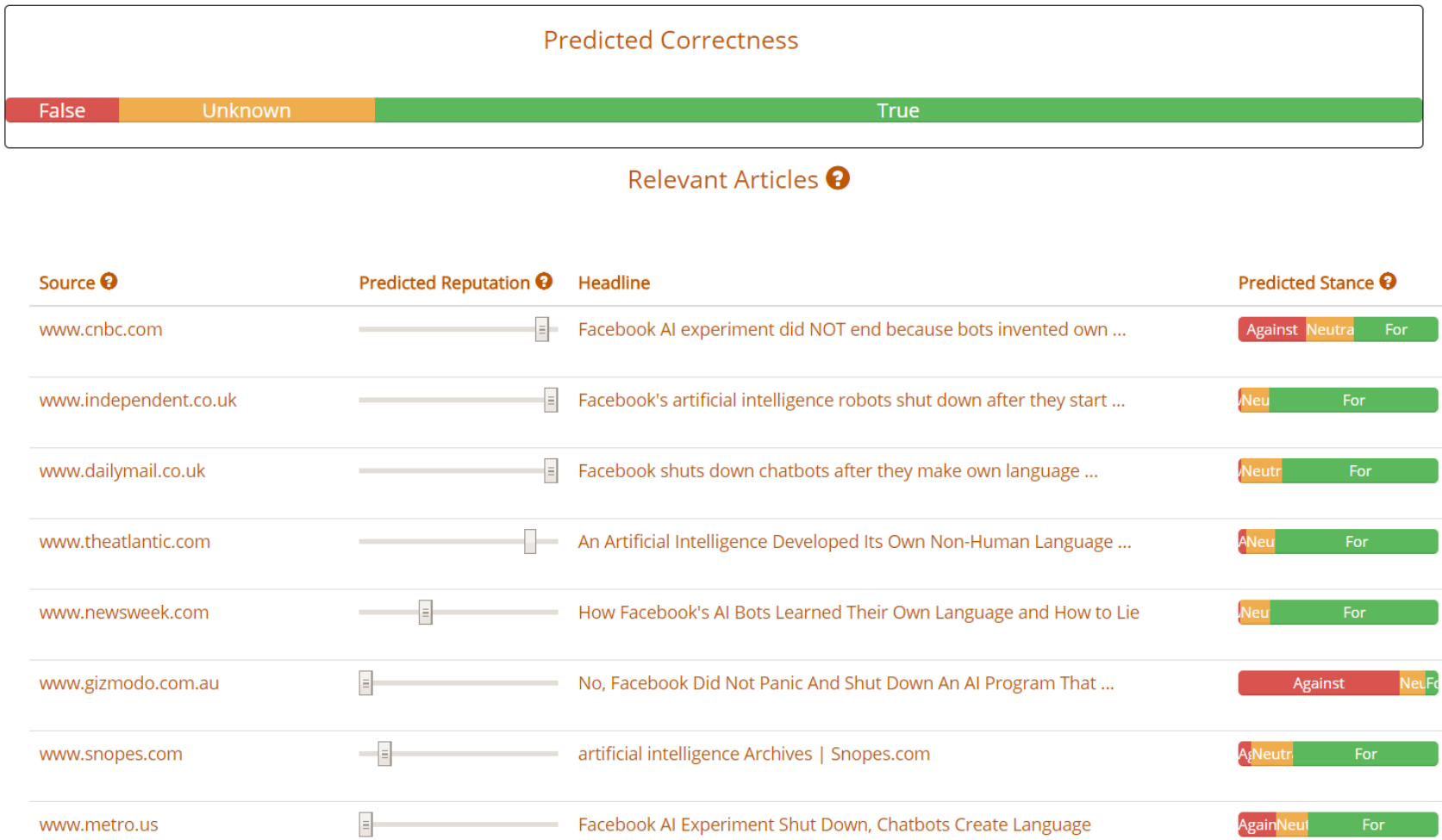
- A: see only veracity predictions
- B: also see explanation (reputation, stances)

User study

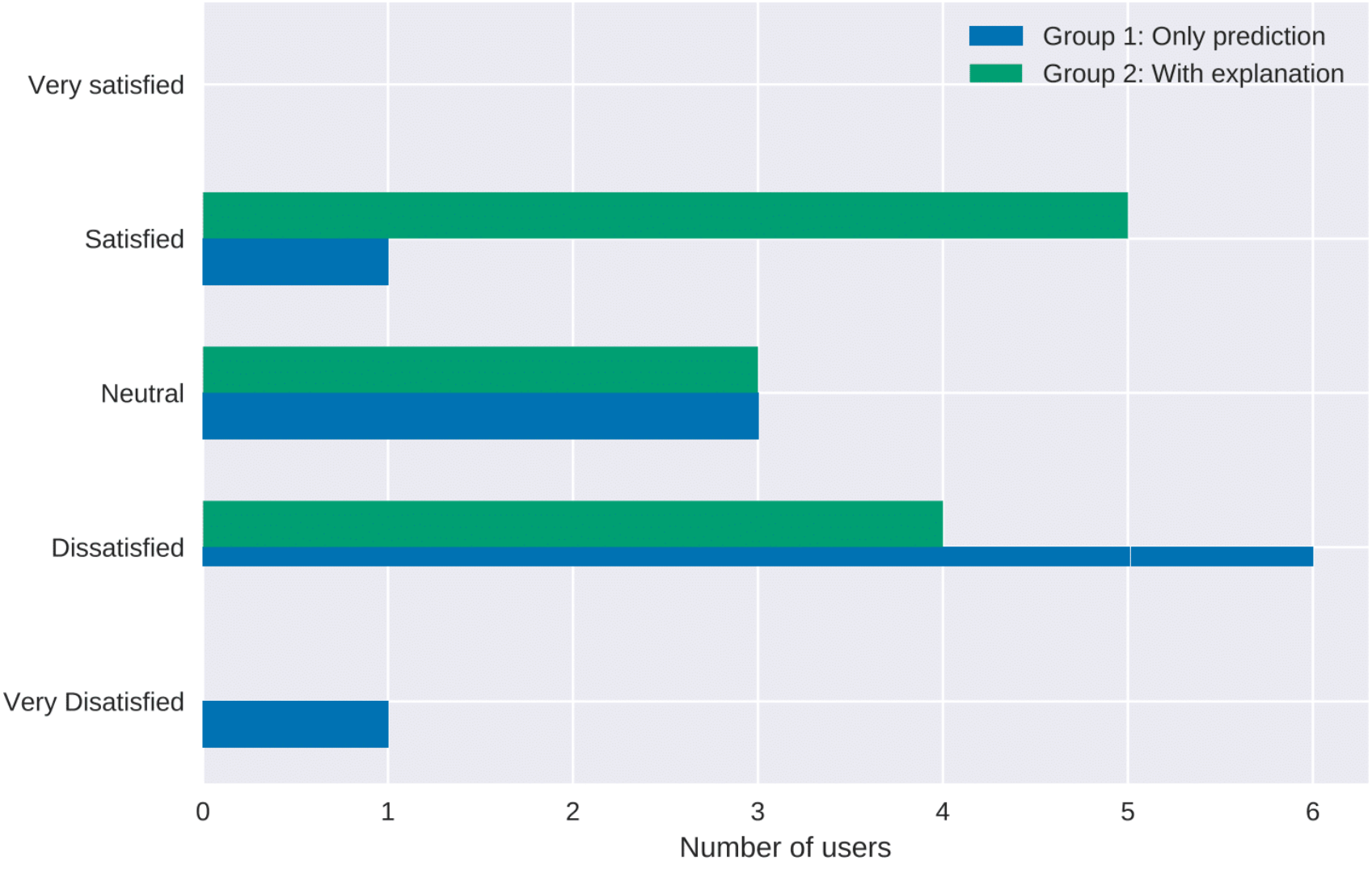
Interface: users enter claims, see predictions.

A/B testing

- A: see only veracity predictions
- B: also see explanation (reputation, stances)



User study: results



Conclusion

Takeaway:

- Stance/Veracity predictions are hard.

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Paper: experiments on Snopes.

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Paper: experiments on Snopes.

Demo: fcweb.pythonanywhere.com

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Paper: experiments on Snopes.

Demo: fcweb.pythonanywhere.com

We share code + data

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Paper: experiments on Snopes.

Demo: fcweb.pythonanywhere.com

We share code + data

Acknowledge: Crowd annotator, reviewers, NSF.

Conclusion

Takeaway:

- Stance/Veracity predictions are hard.
- We contribute: crowdsourcing + joint modeling.

Paper: experiments on Snopes.

Demo: fcweb.pythonanywhere.com

We share code + data

Acknowledge: Crowd annotator, reviewers, NSF.

Questions?