An Annotation Protocol for Collecting User-Generated Counter-Arguments using Crowdsourcing

Paul Reisert¹,³, Gisela Vallejo², Naoya Inoue³,¹, Iryna Gurevych², and Kentaro Inui³,¹
¹RIKEN Advanced Center of Intelligence Project (AIP), Tokyo, Japan
²Ubiquitous Knowledge Processing (UKP), Department of Computer Science, Technische Universität Darmstadt, Darmstadt, Germany
³Graduate School of Information Sciences, Tohoku University, Sendai, Japan

Overview
• Constructive feedback (CF) improves critical thinking skills
• Useful as CF: counter-arguments (CAs) and fallacies
• Little work on CAs for micro-level (one claim/premise) arguments

Challenge
• Collecting large-scale CAs useful for a model is challenging

Solution
• Create an annotation protocol for collecting large-scale CAs

Contribution
• Created protocol for collecting CAs via crowdsourcing
• Conducted qualitative analysis on CA quality

Crowdsourcing
• Useful for collecting large-scale data
• Instructed workers to i) create CA (CAG) and ii) create CA after identifying fallacy (CAG-F)
• Conducted trials using the Argument Reasoning Comprehension dataset (Habernal+, 2018)
• 5 random fallacy types

Large-scale CA corpus
• Created large-scale corpus after submission
• Includes 4 fallacy types

Statistics

<table>
<thead>
<tr>
<th>Criteria</th>
<th>HG</th>
<th>BQ</th>
<th>QC</th>
<th>RH</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsure</td>
<td>2,043</td>
<td>315</td>
<td>2,136</td>
<td>1,879</td>
<td>6,373</td>
</tr>
<tr>
<td>CAs (FS)</td>
<td>3,365</td>
<td>3,818</td>
<td>2,121</td>
<td>1,772</td>
<td>11,076</td>
</tr>
<tr>
<td>CAs (O)</td>
<td>907</td>
<td>2,182</td>
<td>2,058</td>
<td>2,664</td>
<td>7,811</td>
</tr>
<tr>
<td>CAs (total)</td>
<td>4,272</td>
<td>6,000</td>
<td>4,179</td>
<td>4,436</td>
<td>18,887</td>
</tr>
</tbody>
</table>

Contact
• Contact author: Paul Reisert, paul.reisert@riken.jp

Future Work
• Create a model for automatically identifying fallacies and generating informative CAs

Fallacy Type | Definition
---|---
Hasty Generalization (HG) | Someone generally always assumes something based on a few instances.
Begging the Question (BQ) | The truth of the premise is already assumed by the claim.
Questionable Cause (QC) | The cause of an effect is questionable.
Red Herring (RH) | Someone reverts attention away from the original claim by changing the topic.

Topic: Are police too willing to use force?

Argument A1: Police are too willing to use force. Police are using excessive force all over the U.S.

Constructive Feedback

Teacher: The use of force causes less violation of the law. Not all actions of the police are violent. People who talk about police force are people who have been arrested.

Output: Revised Argument R1: Police are too willing to use force, but as a result, crime is reduced. Although many people think that arrested individuals discuss this issue....

Annotation Study
• Annotate quality of collected CAs
• Observe inter-annotator agreement (IAA)
• Annotators: two fluent-English speakers
• 100 instances for each task (CAG & CAG-F)

<table>
<thead>
<tr>
<th>Stage</th>
<th>IAA</th>
<th>#Agreed</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAG</td>
<td>0.29</td>
<td>54/100</td>
</tr>
<tr>
<td>CAG-F</td>
<td>0.37</td>
<td>74/100</td>
</tr>
</tbody>
</table>

Corpus of large-scale CAs

Topic: Immigration really a problem?
Argument: Immigration is really a problem because illegal immigrants crowd the schools which in turn causes higher property taxes.

CA: If "illegal immigrants crowd the schools" is assumed to be true, then "immigration is really a problem" is already assumed to be true. (BQ)

CA: There is a questionable cause in this argument because "illegal immigrants attending school" does/will not cause "higher property taxes". (QC)

CA: if "illegal immigrants crowd the schools" is assumed to be true, then "immigration is really a problem" is already assumed to be true. (BQ)

Worker

Workers