Roles and Success in Wikipedia Talk Pages: Identifying Latent Patterns of Behavior

Keith Maki, Michael Miller Yoder, Yohan Jo, and Carolyn Penstein Rosé
Language Technologies Institute, Carnegie Mellon University
Collaborative Discourse
Collaborative Discourse
Collaborative Discourse
Discourse Analysis

● Extract meaningful aspects of discussion
● Support interventions to improve discourse
  ○ Dialog systems, CSCW, MOOCs, consumer apps
Wikipedia

- Article edits
- Talk page contributions
Article edits

Talk page contributions
Wikipedia

- Article edits
- Talk page contributions
Wikipedia

- Article edits
- Talk page contributions

Definition by Samuel  [edit]

The definition by Arthur Samuel,(1959) seems to be non-existent. Some papers/books cite his key-paper on ML in Checkers-games (see: http://ailtopics.org/sites/default/whichever (better yet, it states "While this is not the place to dwell on the importance of machine-learning procedures, or to discourse on the philosophical aspects" p. where that definition is stated :)

Agree with above - this is a clear problem, as the WP leading quote can be found in many, many places around the Internet (as of 2017) with no actual citation. I've ma 16:17, 14 August 2017 (UTC)

The second source added by User:HelpUsStopSpam is behind a paywall and so isn't clear on the content. Can you excerpt the exact phrase and context used in that ?

Yes, this is a problem that should be solved. Why hasn't it been? The first sentence absolutely does not need to contain the definition from the first time the term oc changed and deepened enormously since 1959. I suggest a paraphrase of this: the difficulties face by systems relying on hard-coded knowledge suggest that AI sy learning. Goodfellow, Bengio, Courville; Deep Learning; MIT Press; 2016; page 2. --Ettrig (talk) 10:43, 13 November 2017 (UTC)

The "definition" paraphrased from Samuel seems to be the the most common one. The second source (Koza et al. 1996) says "Paraphrasing Arthur Samuel": "I paraphrased-from attribution to Arthur Samuel. A) Arthur Samuel is frequently cited/paraphrased throughout literature, this ("without being explicitly programme we've seen here just re-iterate what they read in other works that repeated what they read in other works and so on. Goodfellow and Bengio is certainly not a be networks. I'd rather stick with Arthur Samuel. Chire (talk) 12:24, 13 November 2017 (UTC)
Related work

- Edits play role in shaping relationships between editors (Kittur and Chi, 2007)
- Talk page provides forum for deliberation, information sharing, policy discussion, and off-topic remarks (Viégas and Ham, 2007)
- Level of power impacts roles users play on talk pages (Danescu-Niculescu-Mizil and Kleinberg, 2012)
Related work

- Analyzing and understanding roles Wikipedia users play
  - Edit Behavior (Arazy et al. 2017; Yang et al. 2016)
  - Talk page (Ferschke et al. 2015)
Role Modeling

- Learn latent “roles” played by participants
Role Modeling

- Learn latent “roles” played by participants
- Roles operationalized as patterns of behavior
Role Modeling

- Learn latent “roles” played by participants
- Roles operationalized as patterns of behavior
- Capture functional interplay between discourse participants
Role Identification Model (Yang et al. 2015)

- Assign users in C teams to K roles
- Maximum weighted bipartite matching
- Iteratively update role weights and user assignments

Figure 1: Weighted bipartite graph for matching users and roles
Role Identification Model  (Yang et al. 2015)

Limitations
1. All roles are present in every conversation.
2. Each role is played by exactly one editor.
3. Each editor plays exactly zero or one role(s).
4. All behaviors from an editor represent their role.
5. Behaviors from editors with no role are ignored.
6. Editors independent across conversations
Proposed approach

- Represent roles in conversations using graphical model
- Model user behavior as a mixture of roles
- Relax assumptions e.g. all roles present

Figure 1: PRPM plate diagram relating for each conversation $N$ the outcome measure $y$ and each user $M$’s $L$ behaviors $b$. 
Dataset

- Data extracted using JWPL (Zesch and Gurevych, 2008)
- Sampled discussions from 100k articles alongside their edit histories from 2004 to 2014
- Filter to discussions which have 2 or more participants who also edit the article within 1 week of the discussion

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>number of articles</td>
<td>7,211</td>
</tr>
<tr>
<td>number of discussion threads</td>
<td>21,108</td>
</tr>
<tr>
<td>number of editor-discussion pairs</td>
<td>53,175</td>
</tr>
<tr>
<td>average #editors/discussion</td>
<td>2.52</td>
</tr>
</tbody>
</table>

Table 1: Dataset statistics
Dataset

- Model roles in talk page discussions surrounding shifts in article perspective (Priedhorsky et al. 2007)
- Define editor success score for an editor in a thread as the proportion of tokens changed that remain changed following the discussion

\[ y(u,t) = 1 - \frac{\sum_{i=1}^{n} |c_i|}{\sum_{i=1}^{n} |e_i|} \]

- \( y(u,t) \) - the score for editor \( u \) in discussion thread \( t \)
- \( e_i \) - the associated edits by editor \( u \)
- \( c_i \) - the tokens of \( e_i \) changed by other editors
Experiments

● Regression predicting editor scores based on contextual discussion behavior of editors
  ○ Outcome measure - success score of one editor
  ○ Features
    ■ Dialogue Act Features (Jo et al. 2017)
    ■ Behavior Features
      ● Position of the editor in discussion
      ● Style characteristics (Tan et al. 2016)
      ● Authority claims (Bender et al. 2011)
      ● Emotion expressed (Tausczik and Pennebaker 2010)
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Setting</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5+</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinReg</td>
<td>tgt editor</td>
<td><strong>0.286</strong></td>
<td><strong>0.302</strong></td>
<td><strong>0.287</strong></td>
<td>0.302</td>
<td>0.292</td>
</tr>
<tr>
<td>LinReg</td>
<td>all</td>
<td>0.287</td>
<td><strong>0.302</strong></td>
<td>0.289</td>
<td>0.301</td>
<td>0.292</td>
</tr>
<tr>
<td>RIM</td>
<td>$K=2$</td>
<td>0.316</td>
<td>0.317</td>
<td>0.308</td>
<td>0.342</td>
<td>0.318</td>
</tr>
<tr>
<td>RIM</td>
<td>$K=3$</td>
<td>0.307</td>
<td>0.320</td>
<td>0.310</td>
<td>0.337</td>
<td>0.314</td>
</tr>
<tr>
<td>RIM</td>
<td>$K=4$</td>
<td>0.307</td>
<td>0.314</td>
<td>0.311</td>
<td>0.327</td>
<td>0.311</td>
</tr>
<tr>
<td>RIM</td>
<td>$K=5$</td>
<td>0.309</td>
<td>0.315</td>
<td>0.308</td>
<td>0.321</td>
<td>0.312</td>
</tr>
<tr>
<td>PRPM</td>
<td>$K=2$</td>
<td><strong>0.286</strong></td>
<td><strong>0.302</strong></td>
<td>0.288</td>
<td>0.297</td>
<td>0.292</td>
</tr>
<tr>
<td>PRPM</td>
<td>$K=3$</td>
<td><strong>0.286</strong></td>
<td><strong>0.302</strong></td>
<td>0.288</td>
<td><strong>0.295</strong></td>
<td><strong>0.291</strong></td>
</tr>
<tr>
<td>PRPM</td>
<td>$K=4$</td>
<td><strong>0.286</strong></td>
<td><strong>0.302</strong></td>
<td>0.289</td>
<td><strong>0.295</strong></td>
<td><strong>0.291</strong></td>
</tr>
<tr>
<td>PRPM</td>
<td>$K=5$</td>
<td><strong>0.286</strong></td>
<td><strong>0.302</strong></td>
<td>0.288</td>
<td><strong>0.295</strong></td>
<td><strong>0.291</strong></td>
</tr>
</tbody>
</table>

Table 2: RMSE for baselines and models. Rows are model settings. Scores are reported for different numbers of participants, which are the columns headings. (*LinReg: editor uses only the target editor’s features, and all uses all participants’ features. *RIM* and *PRPM*: $K$ is the number of roles.*)
Analysis of Roles

- Qualitative analysis of 5 learned roles
- Compare behavior distributions as well as tendency towards success alone and in context with other roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Example post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderator</td>
<td>It was requested that this article be renamed but there was no consensus for it be moved.</td>
</tr>
<tr>
<td>Architect</td>
<td>I think a section in the article should be added about this.</td>
</tr>
<tr>
<td>Policy</td>
<td>The article needs more WP:RELIABLE sources.</td>
</tr>
<tr>
<td>Work</td>
<td>The name of the article should be ““Province of Toronto”” because that is the topic of the article.</td>
</tr>
<tr>
<td>Wordsmith</td>
<td>There actually was no serious Entnazifizierung in East Germany.</td>
</tr>
</tbody>
</table>

Table 3: Examples of discussion posts from users in certain learned roles
Conclusion and Future Work

- Detail-oriented roles associated with success in combination with organizational roles
- Multiple participants taking organizational roles can lessen individual editing success
- Future work should address limitation that editors assumed to be self-independent across conversations
- Data available at

https://github.com/michaelmilleryoder/wikipedia-talk-scores


References


Stock images courtesy of Master isolated images, dream designs, and Stuart Miles at FreeDigitalPhotos.net. Clip art used freely under public domain.
Analysis of Roles

Figure 3: Behavior distributions for each role, expressed for each behavior as the number of standard deviations above the mean.
Additional Details

- Discourse act annotations

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRITCOMPL</td>
<td>Information is incomplete or lacks detail</td>
</tr>
<tr>
<td>CRITACC</td>
<td>Lack of accuracy, correctness or neutrality</td>
</tr>
<tr>
<td>CRITLANG</td>
<td>Deficiencies in language and style</td>
</tr>
<tr>
<td>CRITSUIT</td>
<td>Content not suitable for an encyclopedia</td>
</tr>
<tr>
<td>CRITSTRUCT</td>
<td>Deficiencies in structure or visual appearance</td>
</tr>
<tr>
<td>CRITAUTH</td>
<td>Lack of authority</td>
</tr>
<tr>
<td>ACTF</td>
<td>Commitment to action in the future</td>
</tr>
<tr>
<td>ACTP</td>
<td>Report of past action</td>
</tr>
<tr>
<td>REQEDIT</td>
<td>Request for article edit</td>
</tr>
<tr>
<td>REQMAINT</td>
<td>Request for admin or maintenance action</td>
</tr>
<tr>
<td>ATTPPOS</td>
<td>Positive attitude</td>
</tr>
<tr>
<td>ATTNEG</td>
<td>Negative attitude</td>
</tr>
</tbody>
</table>

*Table 1. Discourse Act Annotations for the Article Quality Dataset*