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

WIKIPEDIA

Collaborative Discourse

WIKIPEDIA

stackoverflow

Collaborative Discourse

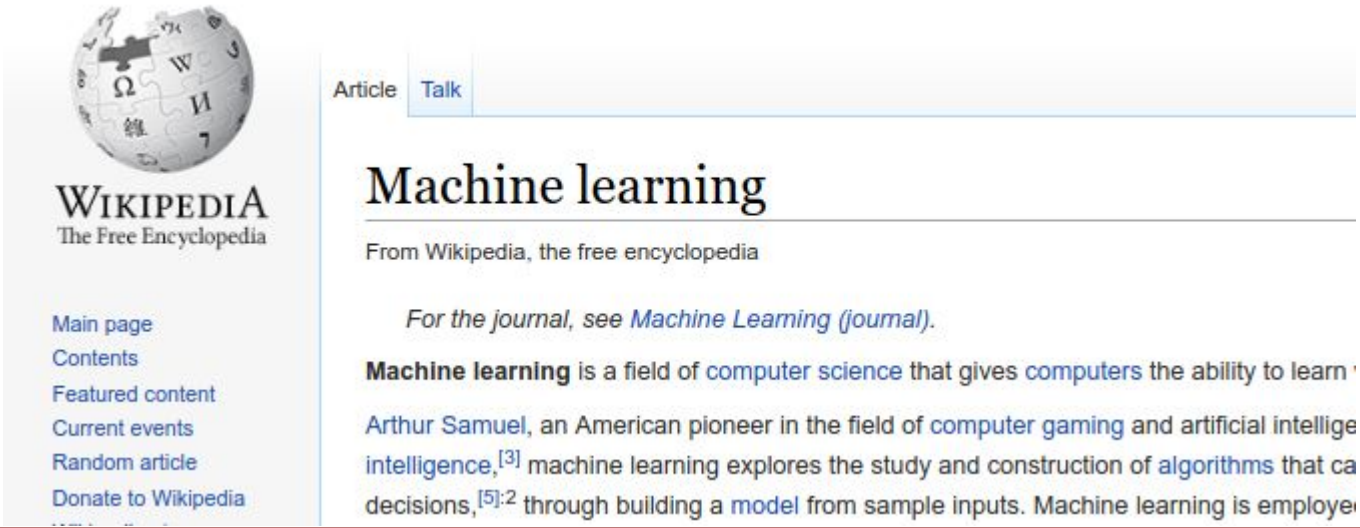
GitHub

WIKIPEDIA

stackoverflow

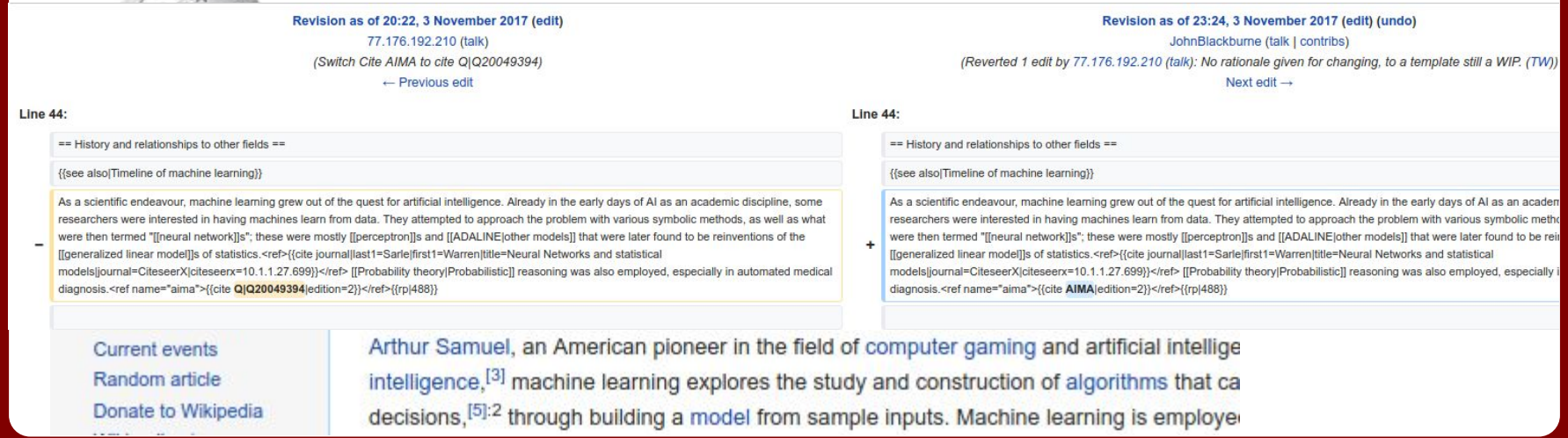
Wikipedia

- Article edits
- Talk page contributions



Wikipedia

- Article edits
- Talk page contributions



Revision as of 20:22, 3 November 2017 (edit)
77.176.192.210 (talk)
(Switch Cite AIMA to cite Q|Q20049394)
← Previous edit

Line 44:

== History and relationships to other fields ==

{{see also|Timeline of machine learning}}

− As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "[[neural network]]s"; these were mostly [[perceptron]]s and [[ADALINE|other models]] that were later found to be reinventions of the [[generalized linear model]]s of statistics.<ref>{{cite journal|last1=Sarle|first1=Warren|title=Neural Networks and statistical models|journal=CiteseerX|citeseerx=10.1.1.27.699}}</ref> [[Probability theory|Probabilistic]] reasoning was also employed, especially in automated medical diagnosis.<ref name="aima">{{cite **Q|Q20049394**|edition=2}}</ref>{{rp|488}}

Revision as of 23:24, 3 November 2017 (edit) (undo)
JohnBlackburne (talk | contribs)
(Reverted 1 edit by 77.176.192.210 (talk): No rationale given for changing, to a template still a WIP. (TW))
Next edit →

Line 44:

== History and relationships to other fields ==

{{see also|Timeline of machine learning}}

+ As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academ researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic metho were then termed "[[neural network]]s"; these were mostly [[perceptron]]s and [[ADALINE|other models]] that were later found to be rei [[generalized linear model]]s of statistics.<ref>{{cite journal|last1=Sarle|first1=Warren|title=Neural Networks and statistical models|journal=CiteseerX|citeseerx=10.1.1.27.699}}</ref> [[Probability theory|Probabilistic]] reasoning was also employed, especially i diagnosis.<ref name="aima">{{cite **AIMA**|edition=2}}</ref>{{rp|488}}

Current events
Random article
Donate to Wikipedia

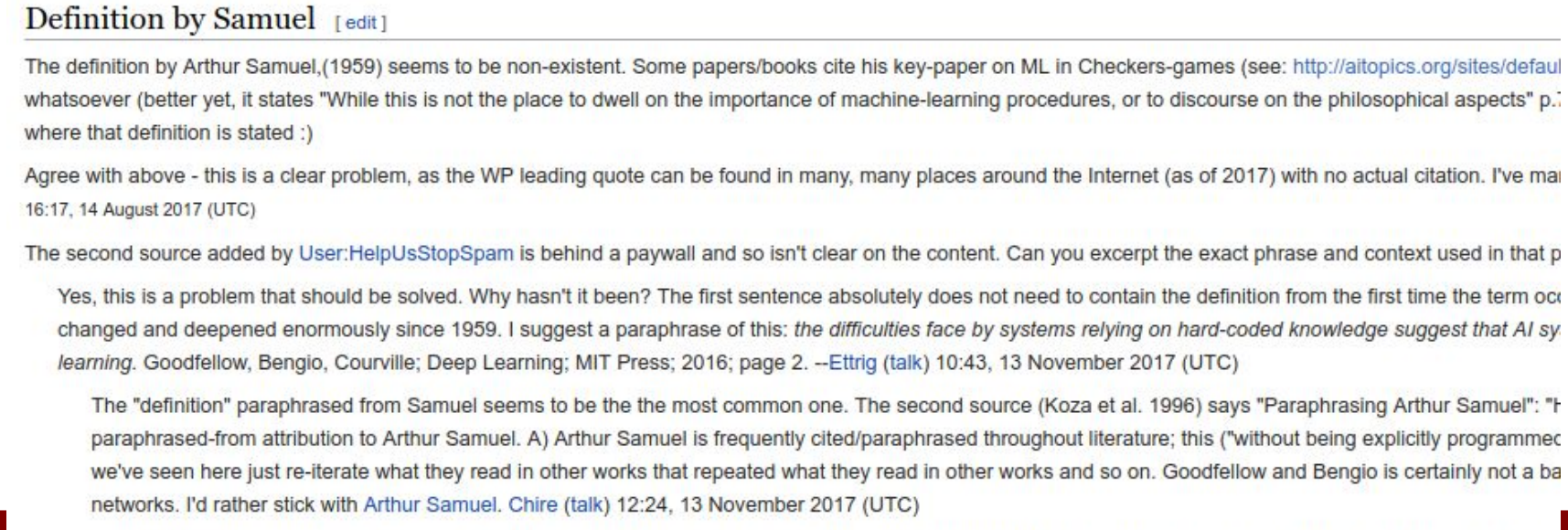
Arthur Samuel, an American pioneer in the field of computer gaming and artificial intellige intelligence,[3] machine learning explores the study and construction of algorithms that ca decisions,[5]:2 through building a model from sample inputs. Machine learning is employe

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://aitopics.org/sites/defaul whatsoever (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.7 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 p

> 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": "H paraphrased-from attribution to Arthur Samuel. A) Arthur Samuel is frequently cited/paraphrased throughout literature; this ("without being explicitly programmed 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 ba networks. I'd rather stick with Arthur Samuel. Chire (talk) 12:24, 13 November 2017 (UTC)

Motivation

- Extract meaningful aspects of discussion
- Understand the nature of discussion about collaborative content
  - How do individuals create influence in collaboration?
  - What effect does discussion have on the content?
  - Relevant to argumentation, CSCW

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

Research Questions

- What discussion strategies are indicative of lasting influence in Wikipedia article edits?

- Are there specific combinations of roles that others take up in discussion that allow individual editors to be more or less successful?

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

number of articles	7,211
number of discussion threads	21,108
number of editor-discussion pairs	53,175
average #editors/discussion	2.52

Table 1: Dataset statistics

Editor Success Score

- Attempt to operationalize influence of an editor's contributions to an article (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

$y(u,t)$ - the score for editor u in discussion thread t

$$y(u,t) = 1 - \frac{\sum_{i=1}^{n} |\mathbf{c}_i|}{\sum_{i=1}^{n} |\mathbf{e}_i|}$$

$\boldsymbol{e}_i$ - the associated edits by editor u

$\boldsymbol{c}_i$ - the tokens of $\boldsymbol{e}_i$ changed by other editors

Example Conversation

Editor	Post (on “Gyro Monorail” page)	Score
GordonVigurs	The mention, in the introduction, that each unit of a monorail train must have a balancing system is no more than a piece of speculation based on intuition, rather than fact.	0.311
DavidWBrooks	Fair enough - I watered it down slightly ("many cars" instead of "all cars"). It's still an obstacle to large-scale adoption.	0.972
PMLawrence	It struck me that a set of two perpendicular Brennan-style stabilisers (allowing freedom of yaw) could make a conventional helicopter inherently stable… Anyway, I was wondering if any work had been done applying this article's systems to helicopters, and if so can anyone supply suitably referenced material about it here?	0.000

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

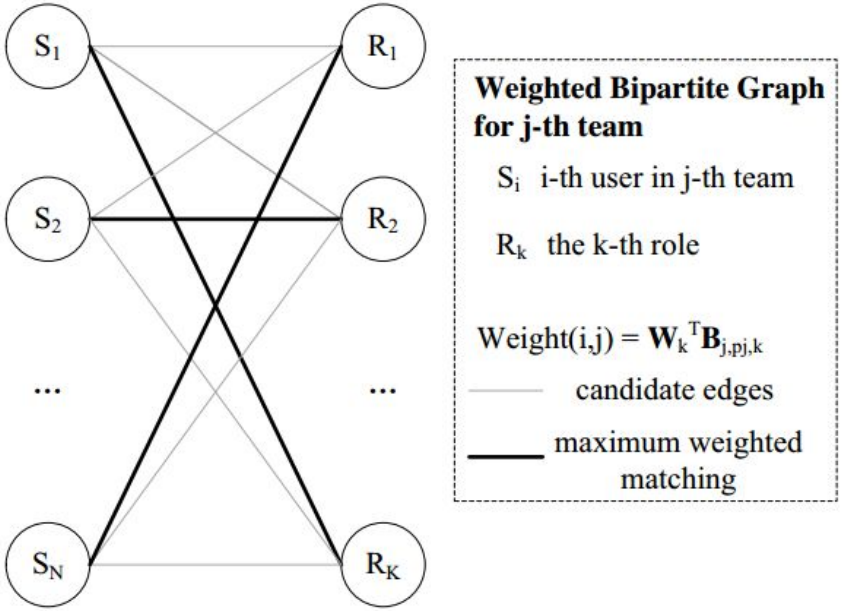


Figure 1: Weighted bipartite graph for matching users and roles

Role Identification Model (Yang et al. 2015)

Assumptions

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.

Probabilistic Role Profiling Model (PRPM)

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

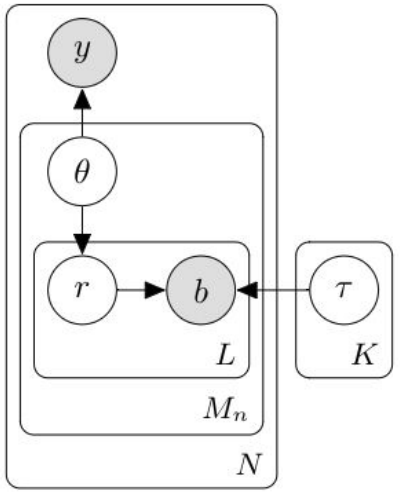


Figure 1: PRPM plate diagram relating for each conversation N the outcome measure y and each user M's L behaviors b.

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

Model	Setting	2	3	4	5+	All
LinReg	tgt editor	**0.286**	**0.302**	**0.287**	0.302	0.292
LinReg	all	0.287	**0.302**	0.289	0.301	0.292
RIM	K=2	0.316	0.317	0.308	0.342	0.318
RIM	K=3	0.307	0.320	0.310	0.337	0.314
RIM	K=4	0.307	0.314	0.311	0.327	0.311
RIM	K=5	0.309	0.315	0.308	0.321	0.312
PRPM	K=2	**0.286**	**0.302**	0.288	0.297	0.292
PRPM	K=3	**0.286**	**0.302**	0.288	**0.295**	**0.291**
PRPM	K=4	**0.286**	**0.302**	0.289	**0.295**	**0.291**
PRPM	K=5	**0.286**	**0.302**	0.288	**0.295**	**0.291**

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

Moderator (low editor success, esp. in groups)

> *"It was requested that this article be renamed but there was no consensus for it to be moved."*

Architect (moderate success)

> *"I think a section of the article should be added about this."*

Policy Wonk (moderately low success)

> *"The article needs more WP:RELIABLE sources."*

Analysis of Roles

Wordsmith (high success,
especially with moderator, architect)

"The name of the article should be "Province of Toronto" because that is the topic of the article."

Expert (moderate success, high in groups)

"There actually was no serious Entnazifizierung in East Germany."

Example Role Assignments

Editor	Post (on “Gyro Monorail” page)	Score	Role
GordonVigurs	The mention, in the introduction, that each unit of a monorail train must have a balancing system is no more than a piece of speculation based on intuition, rather than fact.	0.311	Architect
DavidWBrooks	Fair enough - I watered it down slightly ("many cars" instead of "all cars"). It's still an obstacle to large-scale adoption.	0.972	Wordsmith
PMLawrence	It struck me that a set of two perpendicular Brennan-style stabilisers (allowing freedom of yaw) could make a conventional helicopter inherently stable… Anyway, I was wondering if any work had been done applying this article's systems to helicopters, and if so can anyone supply suitably referenced material about it here?	0.000	Expert/ Architect

Contributions

- Introduce a dataset of Wikipedia talk page conversations paired with associated article edits
- Define a task operationalizing the influence of collaborators in shared content creation
- Present a lightly supervised probabilistic graphical model of discussion roles and behaviors
- Interpret the learned roles to contribute to understanding of collaboration in this context
- Data available at

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

Future Work

- Supervise roles using alternative success score
  - Editor specific
  - Order independent
  - Edits weighted equally
- Further relax modeling assumptions
  - Editors assumed to be self-independent

References

A. Kittur, B. Suh, B. Pendleton and E. Chi. He says, she says: Conflict and coordination in wikipedia. In M. B. Rosson (ed.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 453-462, New York, NY, 2007. ACM.

C. Danescu-Niculescu-Mizil, L. Lee, B. Pang and J. Kleinberg. Echoes of power: Language effects and power differences in social interaction. In M. Rabinovich and S. Staab (eds.), Proceedings of the 21st International Conference on World Wide Web, pp. 699-708, Lyon, France, 2012. ACM.

D. Yang, M. Wen and C. P. Rosé. Weakly supervised role identification in teamwork interactions. In P. Langley (ed.), Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp. 1671-1680, Beijing, China, 2015. ACM.

F. Viegas, M. Wattenberg, J. Kriss and F. Ham. Talk before you type: Coordination in wikipedia. In R. Sprague (ed.), Proceedings of the 40th Hawaii International Conference on System Sciences, pp. 78, Washington, DC, USA, 2007. IEEE.

A. Kittur and R. Kraut. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In B. Begole and D. McDonald (eds.), Proceedings of the 2008 ACM conference on Computer supported cooperative work, pp. 37-46, New York, NY, 2008. ACM.

O. Ferschke, D. Yang and C. P. Rosé. A lightly supervised approach to role identification in wikipedia talk page discussions. In P. Langley (ed.), Proceedings of the 9th International Conference on Web and Social Media Workshop on Wikipedia, Oxford, UK, 2015. AAAI.

R. Priedhorsky, J. Chen, S. Lam K. Panciera L. Terveen and J. Riedl. Creating, destroying, and restoring value in wikipedia. In T. Gross (ed.), Proceedings of the 2007 International ACM Conference on Supporting Group Work, pp. 259-268, New York, NY, 2007. ACM.

S. Chin, W. N. Street, P. Srinivasan and D. Eichmann. Detecting wikipedia vandalism with active learning and statistical language models. In K. Tanaka, X. Zhou, M. Zhang and A. Jatowt (eds.), Proceedings of the 4th Workshop on Information Credibility, pp. 3-10, New York, NY, 2010. ACM.

Analysis of Roles

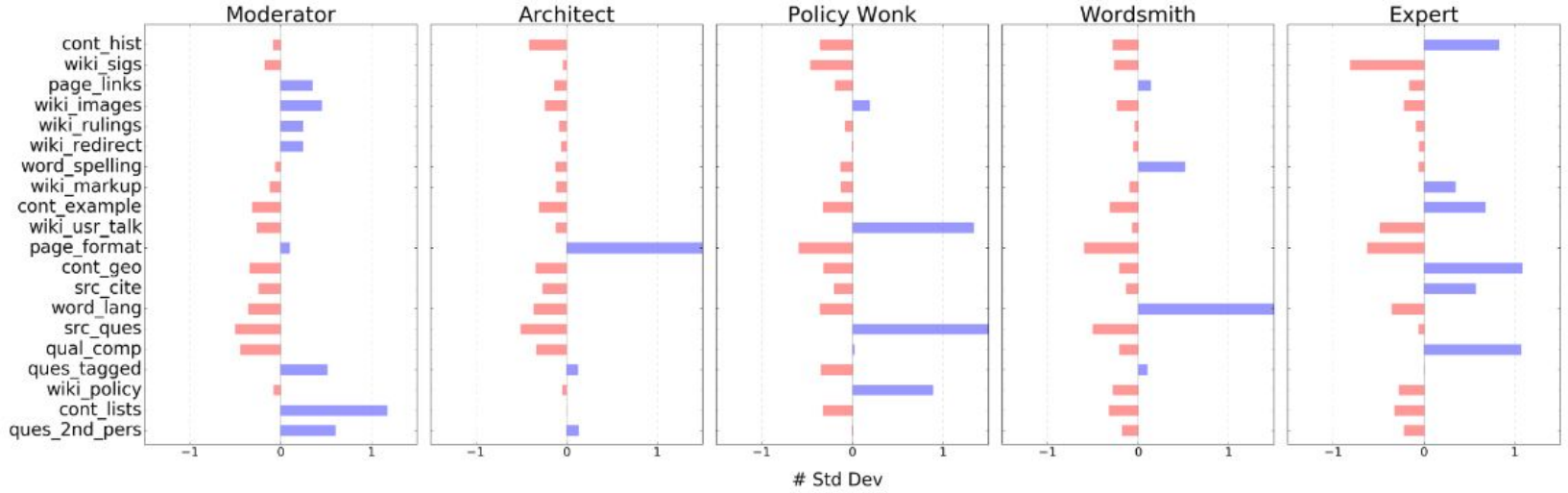
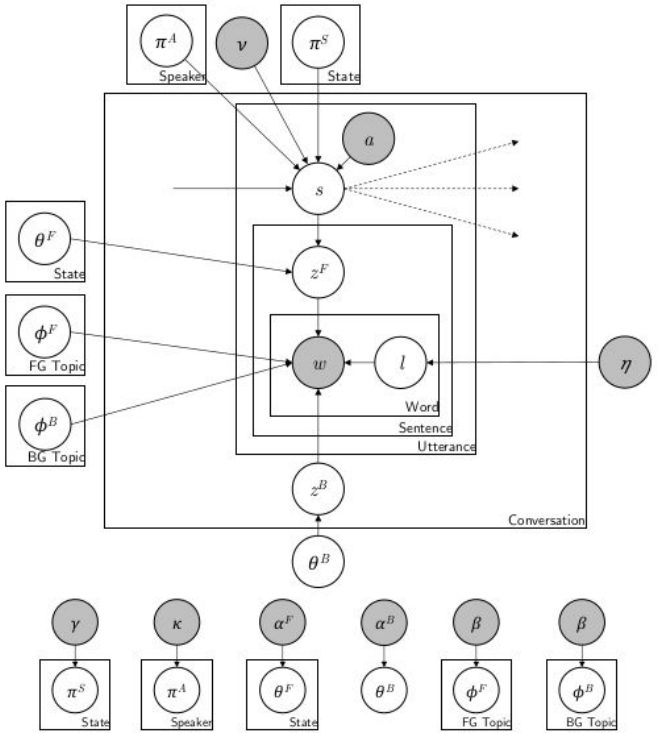


Figure 3: Behavior distributions for each role, expressed for each behavior as the number of standard deviations above the mean.

Additional Details

- Dialogue act features (Jo et al. 2017)
- PGM focusing on content words, speaker preferences



Additional Details

- Argumentation features (Tan et al. 2016)
- definite articles, indefinite articles, positive words, negative words, 2nd person pronoun, links, negative words
- number of common words: $|A \cap O|$,

  reply fraction: $\frac{|A \cap O|}{|A|}$,

  OP fraction: $\frac{|A \cap O|}{|O|}$,

  Jaccard: $\frac{|A \cap O|}{|A \cup O|}$.

Additional Details

- Discourse act annotations

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

Table 1. Discourse Act Annotations for the Article Quality Datset