Supplement for: WikiTalkEdit: A Dataset for modeling Editors' behaviors on Wikipedia

Kokil Jaidka[1]   **Andrea Ceolin** [2]   **Iknoor Singh**[3]
Niyati Chhaya[4]   **Lyle H. Ungar**[5]
`jaidka@nus.edu.sg`

A   Emotion vs Behavior change: data exploration

A.1   Density distribution

In Figure 1a-b, we note that the Editor's comments that trigger positive comments (Mean = 273 characters, Mean Jaccard coefficient, JC = .16) are significantly shorter and have less overlap (content interplay) with the OP's comment than those triggering negative comments (Mean = 417 characters, Mean JC = .18). There was no substantial difference for editorial changes.

Although the density distribution for Fig 1a is somewhat skewed towards shorter comments, a t-test analysis is still valid. For both plots, Welch two-tail t-test between negative emotion and positive emotion has a p-value $< 2.2e\text{-}16$.

A.2   Correlations

Figure 2 presents the major inter-correlations (Spearman's ρ) between the outcomes and the Responder's attributes. Positive correlations are displayed in red and negative correlations in blue color. The color intensity is proportional to the correlation coefficients.

B   General lexical insights

In Figure 3, we can identify the unigrams that are significantly positively (in blue) and negatively (in red) associated with emotional and editorial change are reported after a pairwise Pearson correlation, with Benjamini-Hochberg p-correction. Feedback with a more personal style (first-person pronoun, *thanks*, *hope*) yields a positive emotional change (a). For editorial change, feedback using the second person pronoun (*you*) and evidentiality (*notability/sources/reliable*) is associated with editing behavior (b). However, the use of second person (*you, are*) is associated with negative emotional change, as are questions (*?, what*), negations (*not*), and direct references (*this, that*), cf. (c). The word cloud in (d) does not appear to be very informative but interestingly contains the first person plural pronoun *we*.

In summary, the word clouds show little correlation among features in predicting emotional and editorial change. There seems to be anti-correlation: the use of second-person pronouns is more likely to lead to edits, but also a negative emotional change. The use of first-person pronouns shows the opposite pattern: it is more likely to lead to positive emotional change, but it does not affect editing behavior.

C   Error analysis

More examples of prediction errors are provided in Table 1 and Table 2. Besides the types of errors discussed in the main manuscript, we once again see instances where notability, relevance, and reliability of sources are called into question. Feedback is often implicitly negative, and hard to infer in unsupervised settings.
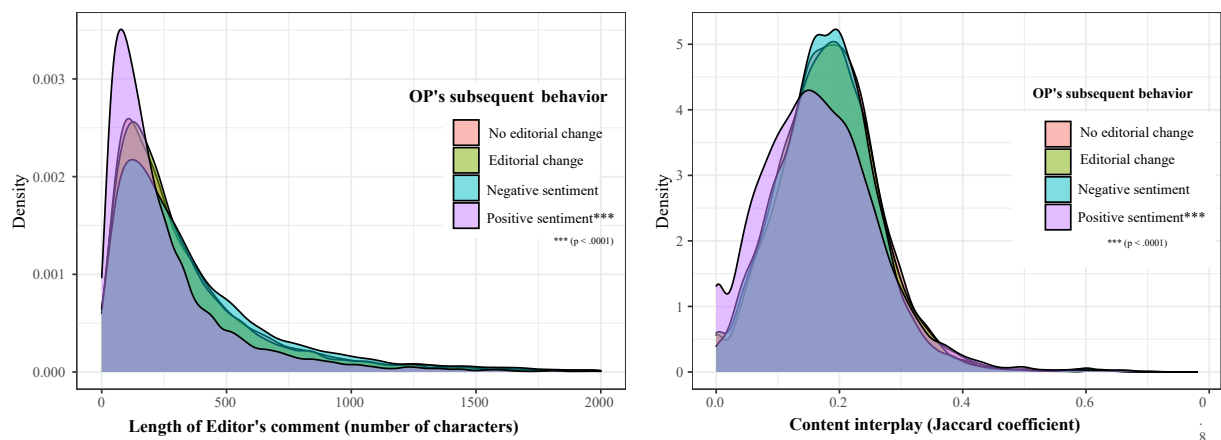
Figure 1: The probability density of sentiment and editorial change as a function of comment length and content interplay.
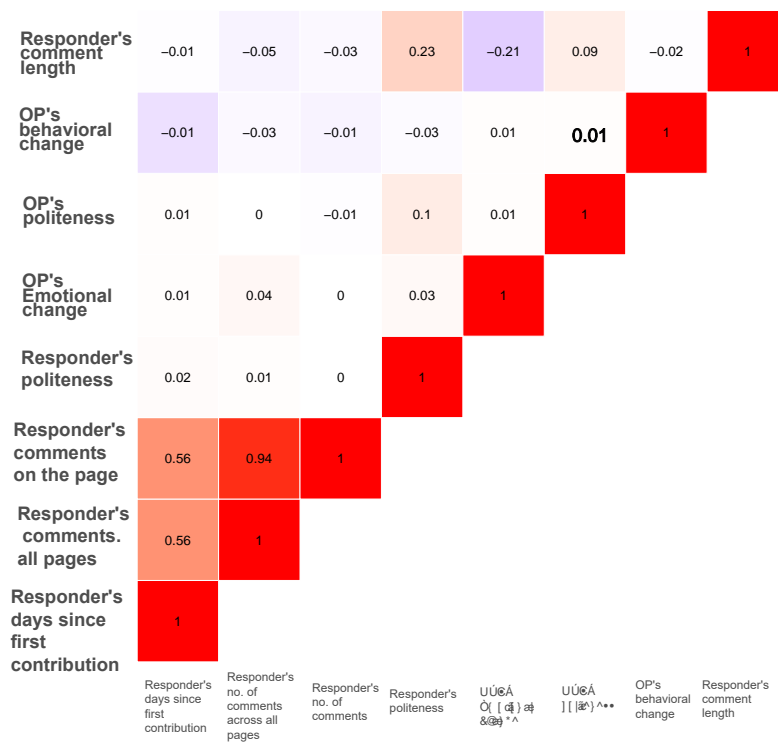

Figure 2: The major inter-correlations (Spearman's ρ) between the outcomes and the Responder's attributes.

(a) Emotion　　　(b) Editing Behavior

Positive change

(c) Emotion　　　(d) Editing Behavior

Negative change

Figure 3: Unigrams visualized in terms of their Pearson correlations with (a) positive emotional change (.01 < r < .05) and (b) editorial change (.03 < r < .05), and (c) negative emotional change (-.05 < r < -.03) and (d) editorial change (-.04 < r < -.02) after Benjamini-Hochberg p-correction (p < 10e-15). The size of the word denotes the magnitude of the correlation while the shade indicates its frequency.

Sentiment change - false positives				
LIWC-based classifier			**XLNet**	
Row. Page	**Talk triplet**		**Row. Page**	**Talk triplet**
Kiev	OP: I am so confused! What is the policy regarding names of towns?[...] E: Danny, there's no real policy. If there is a name which is widely used for one very **famous** city (e.g. Paris, Rome, Athens), then the article with that name should be about that city. [...] OP': Actually, I have a problem with this. Yes, Paris, Rome, and Athens immediately bring to mind great European cities, but why should we play favorites?[...]		Jeff Kennet	OP: ==Massive POV/CRUFT should be deleted== I recommend restoring my edits. This article is filled with POV and cruft. I did not think they would require community consensus but it is what it is. Also, why no mention of the psycho who ǀssaulted him (Fergal Downey)?? Editor: [...] It would have been wiser to bring your concerns to this Talk page, with a little more depth than "This article is filled with POV and cruft" than to so dramatically change the article in one fell swoop. Any chance you could discuss it now? What exactly were your concerns? **What do you want to remove?** What do you want to add? [...] OP': I guess sometimes I am too bold.
2. Aliasing	OP: ...filming a spoked wheel. In my experience, this also happens when watching a spoked wheel. Can someone confirm? [...] Editor: This could be the case if your eye samples the scene - perhaps periferal vision does this. I sometimes get a related effect if I [...] **OP': For really freaky effects**, try watching a TV screen or CRT monitor while using an electric toothbrush. The image wobbles up and down, it was quite **alarming** the first time I saw it.		Religion in Swaziland	OP: ==Improving the article== I have made some edits to improve the article. Editor: **Can you please explain why** you changed CIA statistics? OP': ==Reverted page== I disagree with the edits by , I therefore reverted it to my last edit on July 15.
3. Lesbian, Gay, Bisexual & Transgender Community Center	OP: The webmistress and mods there are really confrontational with new users. They tend to berate you (or ban you outright) if you don't follow their set of rules. [...] Editor: **Hey there!** I see that you're fairly new to Wikipedia, so forgive me if I sound condescending, but I'm just trying to explain a few things. For the sake of coherency, posting the same message in two places isn't the best practice; [...] OP': *What planet are you from? They're not really clever there, unless you enjoy seeing their mods attacking newbies who don't conform to their little in-crowd. [...]		Proxy bomb	OP: ==Retitle== How about retitling this to IRA Proxy bomb? The term Proxy bomb by itself is not very informative. Editor: The tactic was not exclusive of the IRA per cited sources (see lede) OP': All the more reason to denote the particular attacks.
4. Moors	OP: Several days have gone by and [user] has still failed to give me a logical reason for his reversion. I changed the article back to my previous edit and made some [...] Editor: In other words, you have no consensus for the changes. I'm **not surprised** hasn't replied, since you've got further and further from any actual comment on the content of the article. I'm not even sure [...] OP': Don't *revert due solely to no consensus* any of Wikipedia's basic editing policies? This is the second time I have been reverted without a specific reason why.[...]		MetroCard (New York City)	OP: ==Transit Employee ID Cards== This entire subsection is not cited, and I don't think it is accurate. I'd like to add a refimprove section tag [...] Editor: You've ripped the guts out of the article [...] Will someone PLEASE revert this evisceration? Great detail added [...] have been removed. Virtually ALL of which is documented at mta info [...] OP': You're confused.
5. Docking (dog)	OP: == Reverting edit. == My edit got reverted by M, as all my edits usually are. The edit was quite all rigth and was NOT poor quality. I think the reversion was of poor quality. [...] Editor : [user], you present no evidence FOR your edits, only a whine. The puppy photo was too dark to be clear, the boxer photo shows each type of ear clearly. As for the rest, **some may be useful**, but justify the other material you wanted to add.[...] OP': :*The edit was quite all rigth, puppy photo is not too dark to be clear, and calling my conmments to whine is clearly not Wikipedia:Civility. Civility is part of Wikipedia's code of conduct [...]		Li Hongzhi	OP: == Propose to add Li Hongzhi's status as a wanted fellon in China == I would like to add this following fact: [hyperlink] Li is wanted by the Chinese government is a relevant fact ǀand announcement to the effect from the Chinese embassy is a reliable source for this fact.[...] Editor: This information is already in the article but it is sourced to the BBC not the Chinese government Also calling someone a wanted fellon [...] OP': Please cite the relevant wikipedia policy, thanks. Anyways, I'm agreeable to what fellow editors would agree to.

Table 1: Error diagnostics for predicting sentiment change. In this case, our models predicted a positive change ('1'), but the sentiment actually turned negative.

Editorial change - false positives				
LIWC-based classifier			**XLNet**	
Page	**Talk triplet**	**Page**	**Talk triplet**	
Arabic	OP: The redirect to the disambiguation page is intended to assist the reader in locating []...] Editor :Wikipedia is not a dictionary. Adjectives are not supposed to be disamiguated in this way unless they are truly ambiguous. The noun has very clear presedence over the rather nonstandard adjective. See talk:Arab [...] OP': I think you are doing a disservice to readers. I am not going to get in a revert war with you, *but will ask for input from others.*	Mario Kart	OP: == Consoles? == Consoles are non-portable video game systems, handhelds are portable ones. Should we change the uses of Console Games to Video Games due to Mario Kart games on both types? D Editor: Handhelds are really just a type of video game console. (Thus the name of the article we have on it, handheld game console.) Using just "video game" create would instead create a new problem, because arcade game	the main thing we're using to differentiate from, are also a kind of video games. OP': Ah, okay, I thought there was a big difference. Thanks for clearing this up! D [user]	
Oona King	OP :OK, I admit that my own recent edit on this point was not the [...] Editor: *The woman did not appear to be black at all (she looks slightly tanned, almost mediterranean) :Why do you think that being Black is about how a person looks? [...] OP': *There is some merit in your argument, but* race is a matter of genetics, as I see you say yourself. I think we would be a lot better off without racial [...]	Knights Templar legends	OP: == Bannockburn == I've replaced a sentence on the Victorian origin of the myth with a short summary of Cooper's research on Burnes.[...] Editor: Thanks. What really concerned me was [...] We need to cite/attribute this. It's pretty rare to use blogs, see WP:RS and WP:VERIFY. [...] OP': Concern shared. Thanks, we're using the same hymnsheet.	
3. Cricket statistics	OP :== Runs / total innings == There used to be a short paragraph on the suggestion that the batting average should be replaced by [...] Editor: Added back the trivia about the connection between Economy Rate and Clooney. Don't know why this was removed; [...] OP': *I was the one who deleted it.* I didn't think it was relevant to a description about economy rate. I've never come across the term before [...]	Neve Shalom Synagogue	OP: == Requested move == [...] An user changed title Neve Shalom synagogue to Neve Åžalom Synagogue only with his own POV. But it's very clear that the common name in English [...] Editor: If somebody looks at your contribution history they could suspect that you have been following me lately to articles that never ever interested you and try to destroy my honest and objective contributions. OP': Are you serious?	
4. Dread Broadcasting Corporation	OP: Hi there, contacted me through email with the following message. I am reposting it here so it forms part of the discussion on WP of the issues with this article. "I was interested at your constructive contritubutions, and wanted to mention that [...] Editor: I must however question why you have removed the REBEL RADIO stub altogether (which we did not start but changed to address some factual errors in the spirit of Wikipedia, editing not removing) OP': :Hi there. I think this may just be a misunderstanding about how Wikipedia works. Links in this article are pointers to other entries. So making "Rebel Radio" a link here has no impact on [...]	Mimi Smith	OP: ==Mimi== Sorry to do this but I must take issue with you about some of the assertions in Mimi Smith's Wikipedia entry. [...] Perhaps I should start with one which is the easiest to verify. This is the statement that;- âœMenlove Avenue suffered extensive damage during World War IIâ. Nothing of the sort. [...] Editor: Apropos of [...] I have been in contact with Liverpool City Archives and they have not found any particular report about it. [...] I thought that perhaps there might be a reference to [...] Sadly, I never had a reply. OP': Wonderful stuff.	
5. Lords of Conquest	OP: == Restoring an External Link == I owned the site that used to be at http://members.cox.net/[user]. It used to host the Lords of Conquest Imitator, [...] Editor: Hi [user]! I've actually played your version for years... much prettier, faster and better than wheeling out of the ol' C-64. Great job! I've reverted the removal, since it isn't a for-profit [...] OP': Thanks, [user]!	Abraham Jonas (politician)	OP: == Categorization == WP:Categorization of people#General considerations instructs us to "Categorize by those characteristics that make the person notable" - While Jonas was a Freemason [...] Editor: "He then organized the Grand Masonic Lodge of Illinois in 1840 and was elected Grand Master" Seems more than trivial. OP': It might make him note worthy among Masons... but not with the general public. [...] Being a Grand Master is not all that big deal [...]	

Table 2: Error diagnostics for Editorial change. In this case, our models predicted edits ('1'), but no edit was actually made.