Abstract

This supplementary material describes the details of the experiment settings and includes more examples of the experiment results.

3 Data

Topics are extracted using LatentDirichletAllocation in scikit-learn v0.19.1, with the following setting:

- n_components: 20
- max_iter: 200
- learning_method: online
- learning_offset: 50

Generating labels In the CMV forum, DeltaBot replies to an OH’s comment with the confirmation of a $\Delta$, along with the user name to which the OH replied. For most OH replies, the (non-)existence of a $\Delta$ indicates whether a comment to which the OH replied changed the OH’s view. However, an OH’s view is continually influenced as they participate in argumentation, and thus a $\Delta$ given to a comment may not necessarily be attributed to the comment itself. One example is when a comment does not receive a $\Delta$ when the OH reads it for the first time, but the OH comes back and gives it a $\Delta$ after they interact with other comments. In such cases, we may want to give a credit to the comment that actually led the OH to reconsider a previous comment and change the view.

Hence, we use the following labeling that considers the order in which OHs read comments. We treat the (non-)existence of a $\Delta$ in an OH comment as a label for the last comment that the OH read. We reconstruct the order in which the OH reads comments as follows. We assume that when the OH writes a comment, they have read all ancestor comments. So, for each of the OH’s comments in chronological order, all its ancestor comments that have not been considered to be read are attached ...

This ensures that the label of a comment to which the OH replied is the (non-)existence of a $\Delta$ in the OH’s first reply. If an OH reply is not the first reply to a certain comment (as in the scenario mentioned above), or a comment to which the OH replied is missing, the (non-)existence of a $\Delta$ in that reply is assigned to the last comment before the OH wrote it according to our reconstructed order.

5 Experiment

5.1 Model setting

We implemented our model in PyTorch 0.3.0.

5.2 Baseline

We use LogisticRegression in scikit-learn v0.19.1, with the default settings.

5.3 Inputs

TFIDF is extracted using TfidfVectorizer in scikit-learn v0.19.1, with the default setting.

6 Results

Figure 1 and Figure 2 show successful and unsuccessful examples of vulnerable region detection. All examples are from the test set.

Vulnerable region detection Conceptual relevance between a pair of sentences is computed as the cosine similarity between the topic distributions of the sentences.

The first step is to extract topics. Using LatentDirichletAllocation in scikit-learn v0.19.1, we ran LDA on the entire data with 100 topics, taking each post/comment as a document. We treat the top 100 words for each topic as topic words.
this slogan is for people who do not seem to have the iq or common sense to take basic precautions for their own safety. there are two ways to convince these prospective candidates of the darwin award - authority or emotion. appeal to emotion requires some introspection and determining your own worth to your family etc. this is intellectually more involved than common sense and thus clearly beyond the capabilities of these individuals - therefore, an appeal to authority, like law, is your only chance.

but everyone knows there a penalties and fines for breaking the law. its not an appeal to authority, its pointing out the consequences - LRB- the fines - RRB- and appeal to authority would be closer to "buckle up, the government says you should!"
I do not feel obligated to ask permission to take cosplayer pictures at a convention.

I've been to a prominent anime convention ~LRB~ 8000 annual attendees -RRB~ 6 or 7 years now and have never felt the need to ask anyone's permission before taking pictures.

I'll ask permission to take a picture if : * the cosplayer is dressed up as something I really like and no one else is taking their picture - I want them to do their pose or whatever if they do it mind because it's from something I like * they're dressed in something suggestive, showing a lot of skin, or look uncomfortable being dressed that way in a public setting - I do it usually take these people's pictures anyways because 9 times out of 10 me feeling creepy is n't worth the value I'd get having the picture * they might otherwise enjoy being asked to get their picture taken - little girl, something obscure, whatever I typically won't ask to take a picture if : * they've already got a big crowd of people around them taking pictures * they've got a cool costume I want to remember, but I do n't care enough to have them do their pose or whatever .

* I want to capture some aspect of the convention and anime culture itself - to me a convention is like going to a fair or a festival, it's an event I want pictures of, I think the main reason people are so strongly opposed to people taking unwanted pictures is creepy people, and that's a valid concern .

However I think with the general discretion that I follow, asking every single person for their picture is a bit unnecessary.

At the same time, I know a lot of people feel very strongly about photographic consent and I may very well be overlooking something important so change my view!

Edit : wording

---

**OH's initial post**

European style pooping is the worst way to go to the bathroom.

1. Squatting is more comfortable, easier and healthier than sitting.
   - It creates less stress on the the <UNK> muscle allowing for a smoother uninterrupted experience.
   - It plays well with gravity so less pressure is needed and lowers the risk of cancer and other ailments.

2. Toilet paper is messy, expensive and damages the environment.
   - When washed properly the use of your hand is preferable to toilet paper, it might sound disgusting but when you think about it using a thin piece of frail paper to smear around fecal matter with no water or soap is even worse.

3. Modern <UNK> toilets are large, bulky and complex.
   - They take more space, require more maintenance and are ultimately dirtier as butts keep touching them.

---

**Two comments**

* OH's initial post

---

**Comment 1**

I see that as a sort of amateur performance art as someone who has <UNK>, I do n't agree . A street magician, <UNK>, or someone giving a public speech are all asking for your attention, they're doing what they're doing for the sake of their audience. Some cosplayers fit this category, but for some they just want to dress up in a cool costume for the day and a con is the best place to do that.

* OH's initial post

---

**Comment 2**

Would you walk up to someone on the street and take their picture without asking?

---

**Figure 2:** Unsuccessful examples of vulnerable region detection.
<table>
<thead>
<tr>
<th>$n$-grams for $\Delta = 1$</th>
<th>$n$-grams for $\Delta = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>and, in, for, use, it, on, ?, &gt;, sex, why, do_you, wear, relationship, child, are_you, op, mother, should, wearing, teacher, then, it_is, has, deltas, when, same, no, circumcision, you_are, then_you, baby, story</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Top $n$-grams with the most positive/negative weights for logistic regression.

The second step is to compute the topic distribution of each sentence. We simply counted the frequency of occurrences of topic words for each topic, and normalized the frequencies across topics.

Lastly, we computed the cosine similarity between the topic distributions of a pair of sentences.

**TFIDF** The $n$-grams that contribute most to $\Delta$ prediction for logistic regression are shown in Table 1.