SWSR Metadata

Overview

We presents the first effort in creating a hate speech dataset in Chinese. Given the modest presence of Chinese content on Twitter, here we focus on China's prevalent microblogging platform, Sina Weibo. Similar to Twitter, users of Sina Weibo can post new messages (weibos) which can trigger replies from others (comments). Using Sina Weibo to collect weibos and comments related to sexism, we build, annotate and analyse the Sina Weibo Sexism Review (SWSR) dataset. It consists of two parts: *HateWeibo* and *HateComment*, both of which include the textual content of posts along with the information of authors, number of likes and other metadata. The process led to a dataset with 1,527 weibos and 8,969 comments.

In addition, with the aim of assisting research in detection and analysis of sexist comments in Chinese, we provide a sexism-related hate lexicon *SexHateLex* which aggregates and extends existing lexical resources in Chinese.

Data Collection Method

SWSR Dataset Collection

Sina Weibo is the largest microblogging service in China, which has some unique characteristics with respect to Twitter. We use weibo.cn as the source website of Sina Weibo to collect data.

We firstly use keyword search to collect gender-related weibos from Sina Weibo (weibo.cn), where these keywords were chosen based on relevance to the topic and through manual exploration. In addition, we retrieve user profiles and combine these features into the weibo. A total of 9,087 weibos are collected for all keywords in this step.

Then we use weibo ID to extract comments for the collected weibos. This enabled us collection of textual content and metadata of weibos, including user profiles of commenters. This led to the collection of 31,677 comments for the 3,856 weibos.

Finally, we conduct a processing step for collected data (like removing duplicates). This leads to a final set of 8,969 comments from 1,527 weibos for our SWSR dataset.

SexHateLex Lexicon Collection

We aggregate and expand existing Chinese lexicons to build a large Chinese lexicon consisting of terms that can be generally associated with hate speech as well as gender-specific terms. The following lexical resources are referred to:

- Chinese Profanity in Wikipedia https://en.wikipedia.org/wiki/Mandarin_Chinese_profanity\#Sex
- HateBase https://hatebase.org/
- TOCP dataset http://nlp.cse.ntou.edu.tw/resources/TOCP/
- Sexy Lexicon in funNLP https://github.com/fighting41love/funNLP/tree/master/data

Data Structure

SWSR dataset consists of two files: `hateWeibo.csv` and `hateComment.csv`, corresponding to 1527 weibos and 8969 comments. The *SexHateLex* lexicon contains a list of 3016 abusive terms in `SexHateLex.txt`. We only show the format of SWSR dataset here.

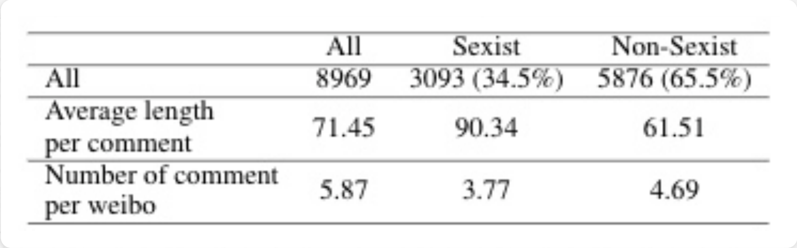
SWSR Structure

- **hateWeibo.csv**
    - weibo_id: a string of weibo ID
    - weibo_text: a string of weibo text
    - keyword: contains the sexist keyword(s) extracted from the weibo. Not every weibo has corresponding keyword(s)
    - user_gender: the gender of user
    - user_location: the location of user
    - user_follower: number of users who follow this user's account
    - user_following: number of users whom this user follows
    - weibo_like: number of like for the weibo
    - weibo_comment: number of like for the weibo
    - weibo_repost: number of like for the weibo
    - weibo_date: the date and time when the weibo is posted
- **hateComment.csv**
    - weibo_id: the weibo id where the comment is collected

- comment_text: a string of comment text
- gender: the gender of commenter
- location: the location of commenter
- like: number of like for this comment
- date: the date and time when the comment is posted
- label: the comment is sexist(1) or non-sexist(0)
- category: categorise sexism into four classes -- stereotype based on appearance(SA), stereotype based on cultural background (SCB), mi-croaggression (MA) and sexual offense (SO)
- target: the type of target who are attacked -- individual (I) or group (G)

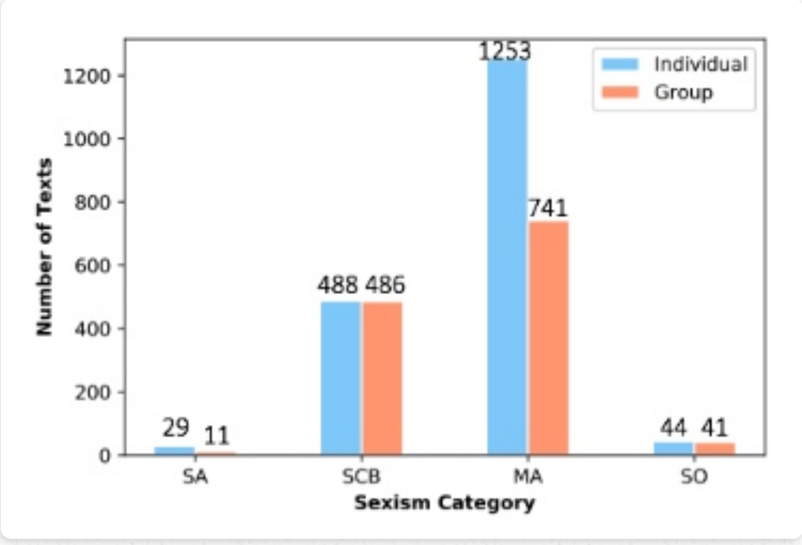
Descriptive statistics

The resulting 8,969 comments are associated with 1,527 weibos. Table below shows the statistics of the dataset in terms of the distribution of sexist comments, comment length and number of comments per weibo.



	All	Sexist	Non-Sexist
All	8969	3093 (34.5%)	5876 (65.5%)
Average length per comment	71.45	90.34	61.51
Number of comment per weibo	5.87	3.77	4.69

The figure below depicts the distribution of the sexism category and target type in sexist comments. More than half of the sexist comments are MA, and SCB also takes a large proportion in the sexist class. Besides, the number of comments towards individuals nearly double those towards groups, where sexist texts in the MA category are more frequently abusive towards individuals.



What's more, we also analyse SWSR dataset and SexHateLex in term of textual distribution, gender distribution and word frequency distribution in dataset paper.

Potential Application

The SWSR dataset can be exploited for building computational methods to identify and investigate online, gender-related abusive language. The SexHateLex lexicon can also support detection and analysis of sexist contents.

Some potential areas of research are:

- Multi-lingual and Cross-lingual Hate Speech Detection
- Cross-domain Hate Speech Detection
- Explainable Hate Speech Detection
- User-based Hate speech Detection
- Other applications with category and target of sexism

More details for these furthering researches will be introduced in the dataset paper.