

X-Posts Explained: Analyzing and Predicting Controversial Contributions in Thematically Diverse Reddit Forums

Anna Guimarães, Gerhard Weikum
Max Planck Institute for Informatics

June 2021



Posted by u/OriginalPoster

9.3k



[Corriere Torino] Sarri last year: "You can send me away, but this team is untrainable"

torino.corriere.it/sport/...

744 Comments **Share**



realreddituser2

Sarri got a ridiculous amount of criticism. The vindication must be nice, good for him.



2.1k



Reply



coolusernamebro

Him being a clown is why we're here. But it's not his fault he was hired, now you guys can see what it's like to have a real manager.

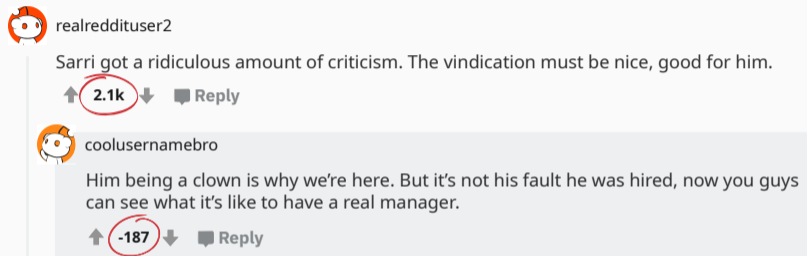


-187



Reply

X-posts and Normal Posts



- Flagged as “controversial” (upvotes/downvotes ≈ 1)
- Score below visibility threshold (upvotes – downvotes < -5)

X-posts explained

- Analyzing and understanding X-posts
- Feature space
 - Activity, sentiments, content topics, cohesion
- Prediction of future X-posts in ongoing discussions
 - Logistic regression classifier

Reddit Datasets

- Subreddits: Politics, World News, Relationships, Soccer (2016-2017)
- Threads (direct-reply post chains) with at least 5 posts



r/politics

1.8M posts

281k X-posts



r/worldnews

1.6M posts

283k X-posts



r/relationships

723k posts

81k X-posts



r/soccer

1.2k posts

107k X-posts

Annotated datasets available at bit.ly/xposts

Discussion Features

- Activity: existing X-posts, number of replies, distinct users



realreddituser2

Sarri got a ridiculous amount of criticism.
The vindication must be nice, good for him.



2.1k



Reply



coolusernamebro

Him being a clown is why we're here.
Now you guys can see what it's like to
have a real manager.



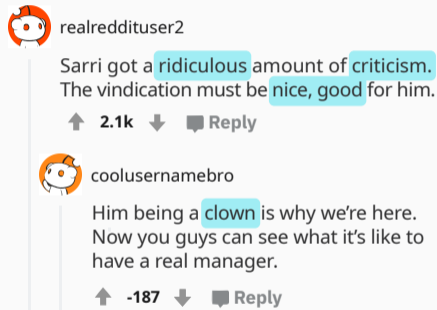
-187



Reply

Discussion Features

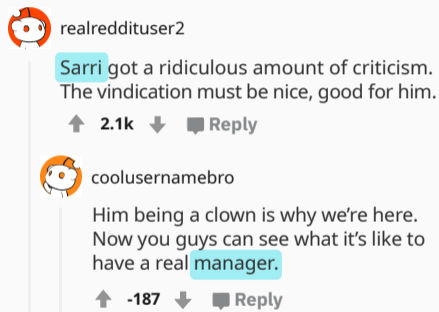
- Activity: existing X-posts, number of replies, distinct users
- Sentiment: polarity shifts and intensity
 - VADER: Valence Aware Dictionary and sEntiment Reasoner¹



¹Hutto, C.J., Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. ICWSM'14.

Discussion Features

- Activity: existing X-posts, number of replies, distinct users
- Sentiment: polarity shifts and intensity
 - VADER: Valence Aware Dictionary and sEntiment Reasoner¹
- Topics: frequent entities and cohesion
 - NER using AIDA²
 - Cosine similarity between posts via sentence embeddings using Doc2VecC³



²Hoffart et al. (2011). Robust Disambiguation of Named Entities in Text. EMNLP'11.

³Chen, M. (2017). Efficient Vector Representation for Documents through Corruption. ICLR'17.

Predicting X-posts in Ongoing Discussions

- Given the features of a discussion's prefix, will its suffix contain an X-post?
 - Logistic regression classifiers using the first four posts in a thread

Source	Year	F1-score	AUC
Politics	2016	0.73	0.79
	2017	0.75	0.77
World News	2016	0.68	0.72
	2017	0.70	0.73
Relationships	2016	0.73	0.79
	2017	0.75	0.80
Soccer	2016	0.67	0.74
	2017	0.65	0.71

Predicting X-posts in the Long-Term

- Classifiers trained on 2016 data and applied to 2017 data
 - Comparable performance to short-term

Source	Year	F1-score	AUC
Politics	2017	0.75	0.80
World News	2017	0.70	0.73
Relationships	2017	0.74	0.80
Soccer	2017	0.64	0.71

Discussion Features

- Informative features in all communities
 - Early X-posts lead to future X-posts
 - Topic cohesion does not!
- Community-specific features
 - Polarizing football teams and players
 - Neutral and negative tone in support-oriented discussions
 - Sentiment shifts and arguments



realreddituser2

There's plenty of less than intelligent people that say Ronaldo is the best of all time

↑ 36 ↓ Reply



coolusernamebro

Messi and Ronaldo are the two best players of all time so thats not wrong

↑ -5 ↓ Reply

Summary

- Short and long-term prediction of X-posts
- Framework for analysis of controversial discussions
 - Additional or alternative features
- Application in content moderation and analysis

Summary

- Short and long-term prediction of X-posts
- Framework for analysis of controversial discussions
 - Additional or alternative features
- Application in content moderation and analysis

Thank You!

Anna Guimarães

aguimara@mpi-inf.mpg.de

