Detection and Analysis of Self-Disclosure in Online News Commentaries

Prasanna Umar, Anna Squicciarini, Sarah Rajtmajer

Introduction

- Users can now share information, express opinions, and discuss various topics of interest with a wide audience online -> privacy risk
- The paper aims to study self-disclosure as it occurs in newspaper comment forums using an automated self-disclosure detector.

Hypothesis

H1: Anonymous users are more likely to self-disclose than identifiable users in online public commentaries.

- Previous work have found that anonymity contributes to increased self-disclosure both on and offline.

H2: Users’ self-disclosure varies across topics.

- Because individuals relate to content based on prior experiences and feelings, they respond differently to different topics.

Related work

- Caliskan et al. [1] created a supervised machine learning method to detect private information in tweets through the use of privacy ontology, named entity recognition, topic modeling and sentiment analysis. -> requires labeled data
- Bak et al. [2] created a supervised method with topic models and SVM to detect personally identifiable information (PII) and personally embarrassing information (PEI.) -> low accuracy
- Bak et al. [3] applied modified latent Dirichlet allocation (LDA) topic models for semi-supervised classification of Twitter conversations into three self-disclosure levels. -> did not categorize texts into individual categories of self-disclosure

Research Design

1. Dataset Collection

- Original dataset consists of 309,319 comments on 52,260 news articles crawled from 10 selected news websites over a three month period.
- After removing duplicates and comments with no readable text, there are about 59,249 comments from 22,132 users (14219 identifiable and 7913 anonymous).

Research Design

Language	Categories	Description
Objective	Birthday/Age	Sharing one's own birthday information or references to own age.
	Race	Sharing one's own race such as being black, white, Hispanic, etc
	Sexual Orientation	Sharing one's sexual orientation and identity such as being straight or LGBT. Includes marital status
	Location	Sharing one's own location such as town, city, states, proximity to a landmark, etc
	Affiliation	Sharing one's nationality, religion, political affiliation, loyalties to groups and brands of a certain nationality, etc
	Money	Sharing one's own financial worth, monetary values of property, plans/goals related to money, etc
	Relationships	Sharing information about the family composition such as having children, brothers, sisters, etc
	Experience	Sharing past experiences of events, habits, work-life, etc. Includes positive or negative experiences and recollections of any past events or memories
Subjective	Interests	Sharing one's own hobbies and interests, including pastimes, favorites, tastes in music, movies, and books. Includes disclosures about pets, as well.
	Feelings	Expressions of deep personal feelings, including humiliation, desires, anxiety, depression, fears, pain, and beliefs which most people would likely disclose only to a friend or family
	Opinion	Discussing one's own opinions, attitudes, and beliefs about current and/or historical events that one is NOT relating to personal experience. Includes views on government, trends, specific events in entertainment/sports, religion, etc.
-	Other	Personal information about the author is revealed but can't be categorized in any categories
-	None	No information about the author revealed

Research Design

2. Self-disclosure detection

- an unsupervised method to detect individual self-disclosure categories in texts.
- Presence of first person pronouns (i, me, my, myself, we, us, our, themselves) is generally considered as linguistic markers for self-disclosing texts.

  (e.g. I live in Pennsylvania / My birthday is in July)
- (1) dictionary construction, (2) subject-verb-object triplet detection, (3) named entity recognition and (4) rule based matching

Research Design

2.1.Category specific dictionary construction

- Dictionary consists of verbs used in a existing Airbnb profiles dataset (interest, lifeMottoValues, work/education, relationship, personality, original residence, travel, hospitality, other)
- extracted the frequent root verbs from sentences in each category to construct dictionaries of verbs.
- For those categories that are not included in Airbnb dataset, the authors manually created additional dictionaries.

Research Design

2.2.Subject,verb, and object extraction

- Used the Python package Spacy for subject, verb and object extraction.
- The subject, verb and object extractor takes text as input, pre-processes the text, splits it into individual sentences and further splits a sentence into individual clauses.

Research Design

2.3.Named entity recognition

- Used the Python package Spacy to see the presence of "real-world object" called a named entity like a location, nationality, etc in a sentence as a distinguishing feature among different categories of self-disclosure.

  (e.g.Date/Cardinal -> birthday/age,

  GPE/FAC/LOC -> location)

Editable Code spaCy v2.2.0 · Python 3 · via Binder

```python
import spacy

nlp = spacy.load("en_core_web_sm")
doc = nlp("Apple is looking at buying U.K. startup for $1 billion")

for ent in doc.ents:
    print(ent.text, ent.start_char, ent.end_char, ent.label_)
```

RUN

```
Apple 0 5 ORG
U.K. 27 31 GPE
$1 billion 44 54 MONEY
```

Research Design

2.4.Rule-based matching

- Objective categories : presence of first-person pronouns + specific category related verb + appropriate named entities
- Subjective categories: presence of first-person pronouns + specific category related verb
- A proximity window of up to 5 words on either side of verb to remove false positives.

  (e.g. I have countless arguments with seemingly educated people in many countries on why Singapore works)

Evaluation

Category (Support)	Precision	Recall	F1-Score
Birthday/Age (37)	19	43	26
Race (21)	46	62	53
Sexual Orientation (13)	50	62	55
Location (188)	26	70	38
Affiliation (106)	27	42	33
Money (150)	50	23	31
Relationships (206)	36	82	50
Experience (464)	29	50	37
None (97)	15	28	19
Subjective: interests, opinions and feelings (2383)	76	73	74
Overall self-disclosure	98	89	93

Hypothesis testing

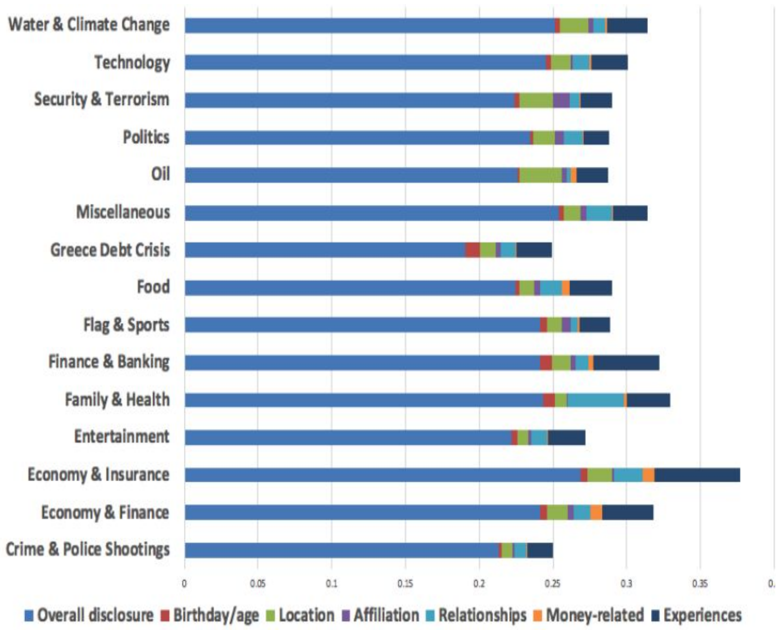
H1: Anonymity and self-disclosure

- aggregated comments by user, obtaining histories of comments for 12, 936 users (5218 anonymous and 7718 identifiable) and labeled users who self- disclosed in at least one comment self-disclosing.
- Chi-squared test and binary logistic regression showed that anonymous users were more likely to self-disclose than identifiable users.

Hypothesis testing

H2: Topics and Self-disclosure

- Ran topic modeling using LDA Mallet anc extracted 15 topics that has the best coherence score.
- ANOVA Results showed that there was a statistically significant difference ir proportion of self-disclosing users between the different topics of discussion.



Thanks