

Understanding The Role Played By Gender In Book Reviews:

A case study of online-reviews on Goodreads

Vaibhav Kumar, Pratyusha Maiti

Do book reviews written by females and males differ in stylistic features?

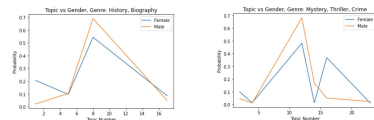
Method: Forward stepwise logistic regression modelling.
Verdict: No strong independent indicators (in contrast to [1]), female-authored reviews characterised by use of pronouns.

Style vs Gender	Mystery, Thriller, Crime	Fantasy, Paranormal	History, Biography
Complexity	1.00232	1.120660	1.039133
Hedging	0.993438	1.011214	1.011214
Lexical Markers	1.004902	1.004075	1.004075
Vocab Richness	0.999945	0.997982	0.999986
Pronouns	0.939842	0.955643	0.954529

Odds Ratio for Stylistic Features vs Gender across genres

Do book reviews written by females and males differ w.r.t topics?

Method: Latent Dirichlet Allocation on review text.
Verdict: Male-authored reviews characterized by topics containing more content-words (like 'book', 'story', 'plot', 'characters').



What is the relationship between the provided rating and gender?

Method: Compare two logistic regression models (with and without gender as i.v) with rating as d.v using ANOVA test (same as [2]).
Verdict: The relationship varies across genres.

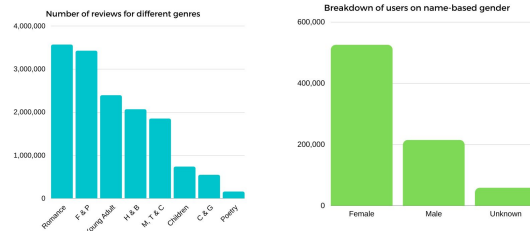
Genre	M, T & C	Children	H & B	Poetry	Comics	F & P
P-value	0.51	0.20	6e-4	0.22	4e-4	4e-3

Objective

Investigate the relationship between gender and online book reviews with respect to perceived utility, sentiment, stylistic features and content.

Data

Reviews: UCSD Book-graph dataset
Public user-info: Goodreads API



Name-based gender

Use first name to identify the gender using [Name2GAN*](#) (Based on collection of names from internet communities)

Labelling-test on names which include self-reported genders in the dataset indicate an F-1 score>0.9.

Do book reviews written by females and males differ w.r.t their content ?

Method: Forward stepwise logistic regression modelling.
Verdict: Entropy is the strongest indicator of male-authored reviews (in keeping with results from [1]).

Content vs Gender	Mystery, Thriller, Crime	Fantasy, Paranormal	History, Biography
Entropy	1.461543	1.01560	1.01560
Brunets	0.790936	0.958956	0.958956
Simpson	0.689575	0.956655	0.956655

Odds Ratio for Content Features vs Gender across genres

Are book reviews written by males perceived as more useful than those written by females ?

Method: Logistic regression modelling.
Verdict: Current evidence indicates no statistically significant difference. In contrast to results from [1].

Ethics*

A current limitation of our study is the binary treatment of gender and a name-based heuristic for determining gender. The specifics of the study have been submitted to the IRB for approval and are under review at the time of writing.

References

1. Otterbacher, Jahna. "Gender, writing and ranking in review forums: a case study of the IMDb." Knowledge and information systems 35.3 (2013)
2. Noei, Ehsan, and Kelly Lyons. "A study of gender in user reviews on the Google Play Store." Empirical Software Engineering 27.2 (2022): 1-28.