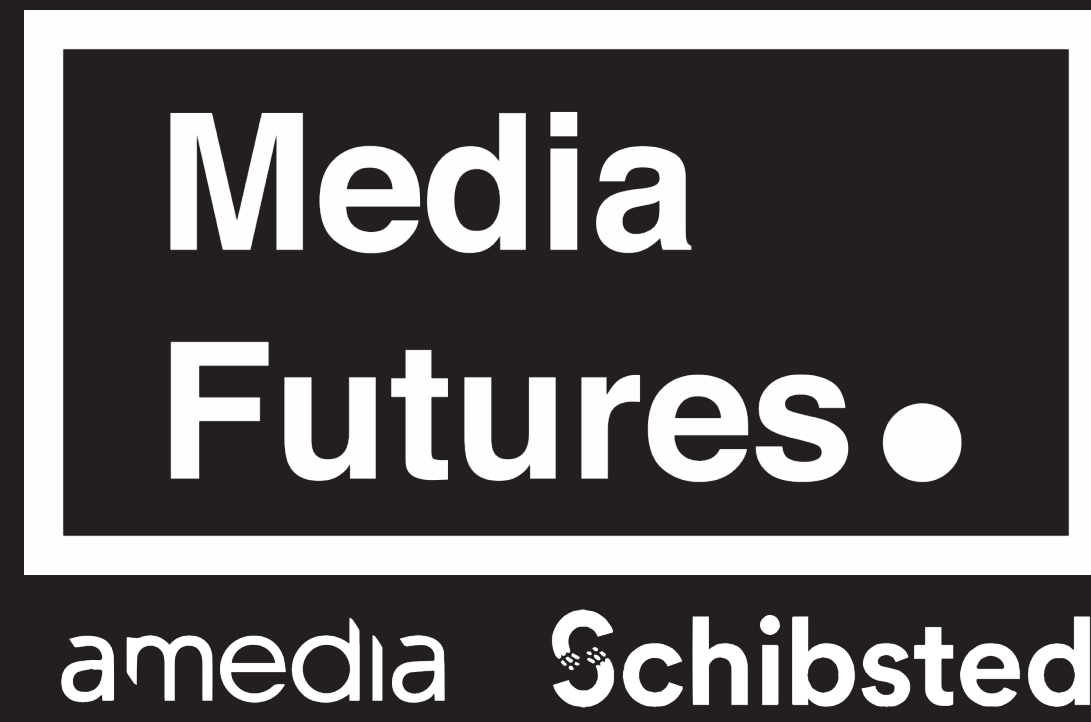


Evaluating Feature-Specific Similarity Metrics using Human Judgments for Norwegian News

Daniel Rosnes

Supervisors: Christoph Trattner and Alain D. Starke



Section: Publisert

Title: Large protests against police violence again fill streets of U.S. cities

Main image:

Author: By David A. Finkelstein

Date of Publication: June 3, 2020 09:23 av: GPT-2

Lead paragraph: Large protests against police violence again filled the streets of several U.S. cities on Tuesday, as elected leaders struggled to cope with the damage already done — citizens injured by police, police tear by civilians and looting who used the protests as cover.

Body text: One of the largest peaceful protests on Tuesday was in Houston, hometown of George Floyd, the African American man whose death in Minneapolis police custody ignited an unprecedented national wave of marches and demonstrations. Journalists on the scene estimated there were 25,000 members, including Houston Mayor Sylvester Turner (D) among them. (AP Photo/Chris Wedel)

Item recommendations:

- 1 Rep. Steve King, who was shunned by GOP leaders for his racist remarks, loses in Iowa primary
- 2 Inside the push to tear-gas protesters ahead of a Trump photo op
- 3 Barr personally ordered removal of protesters near White House, leading to use of force against largely peaceful crowd
- 4 Trump's naked use of religion as a political tool draws rebukes from some faith leaders
- 5 Analysis: What is antifa?

News Features and list of recommendations

Abstract

This master's thesis delves into the measurement of similarity between news articles within the Norwegian news domain. Four central questions form the basis of the thesis: the identification of information cues utilized by readers, the effectiveness of specific similarity metrics, the comparison with other domains, and the exploration of differences in human similarity ratings between national and local news.

Key findings include that a Sentence-BERT metric, applied to the body text, best represented human similarity judgments. Compared to other news domains, the Norwegian news domain showed stronger correlations for a majority of the metrics.

A minimal contrast was observed between human ratings for local and national news, with local news considered slightly more similar. This disparity between local and national levels, however, did not markedly impact how metrics represented human similarity judgments. The findings from this thesis may provide valuable insights for enhancing news recommendation systems within the news sector.

Research questions

- RQ1:** Which specific features do Norwegian users use to evaluate the similarity between news articles?
- RQ2:** To what extent do feature-specific similarity metrics represent human similarity judgments in the Norwegian news domain?
- RQ3:** How do the feature-specific metric representations in the Norwegian news domain compare with feature-specific metric representations in other domains?
- RQ4.1:** Does the strength of human similarity judgments towards Norwegian news media differ across local and national outlets?
- RQ4.2:** To what extent do feature-specific similarity functions represent human judgment across local and national Norwegian news media outlets?

Methodology

Four datasets were generated from two MediaFutures partners: Amedia and Schibsted. The datasets represented all articles published online in 2022 from BA, BT, VG and Nettavisen. Within each dataset, similarity was calculated using 25 different Feature-Specific Similarity Metrics.

For the survey, the similarity scores were normalized and combined and pairs were generated from the combined similarity scores. These pairs were split into 5 groups based on their similarity. 141 participants were recruited through social media, and then asked to rate the similarity of 10 news article pairs on a 5-point Likert scale. Each user rated 5 pairs from a local publication and 5 pairs from a national publication. From each publication, 1 pair was from each of the similarity groups. Finally, each user was asked to report which feature they used to evaluate the similarity.

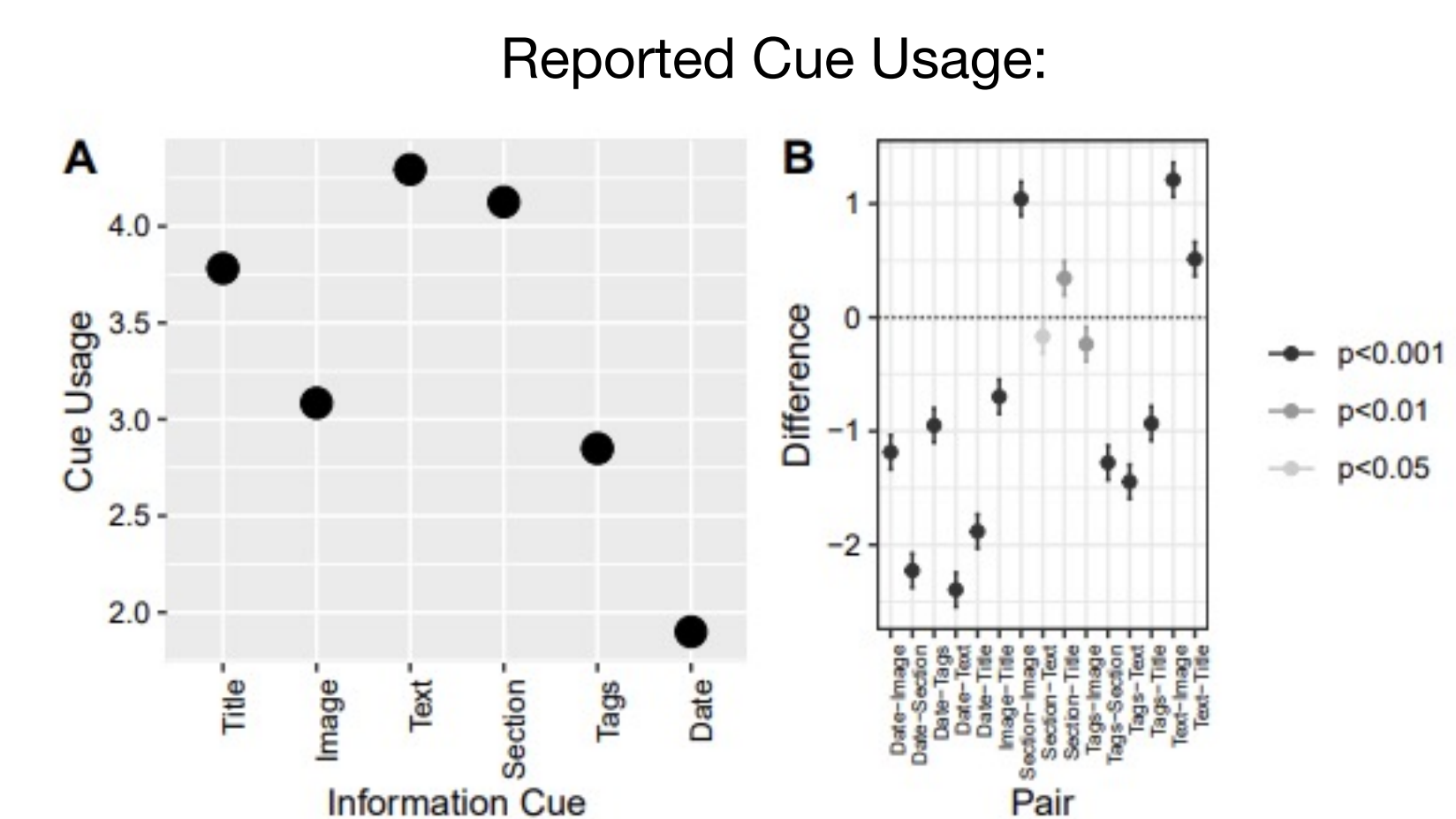
Correlations were calculated for each of the 25 Feature-Specific Similarity Ratings against the Human Similarity Judgements. In addition a t-test was performed to evaluate differences in similarity ratings across Local and National publications. Finally, the correlations were evaluated across the Local and National news domains using a z-test.

Results

- Text and Section highest reported cue usage
- Text:SBERT highest correlating Feature-Specific Similarity Feature, followed by Section:JACC
- Minor differences in similarity ratings across news domains
- Little to no difference in strength of different Feature-Specific Similarity Metrics across National and Local news domains.

Spearman correlations of Feature-Specific Similarity Metrics vs Human Similarity Judgements:

Metric	Publication					
	All	National	Local	VG	BT	Nettavisen BA
Image:BR	0.24***	0.16***	0.32***	0.06	0.36***	0.26***
Image:SH	0.26***	0.24***	0.28***	0.08	0.28***	0.40***
Image:CO	0.13***	0.11*	0.15***	0.12*	0.15*	0.10
Image:COL	0.07*	0.07	0.08	0.11	0.11	0.05
Image:EN	0.22***	0.15***	0.28***	0.09	0.29***	0.21***
Image:EMB	0.30***	0.39***	0.23***	0.32***	0.20***	0.46***
Text:BERTopic	0.40***	0.42***	0.37***	0.39***	0.36***	0.46***
Text:LDA	0.29***	0.29***	0.29***	0.34***	0.33***	0.29***
Text:NENTS	0.21***	0.22***	0.2***	0.12*	0.27***	0.36***
Text:SBERT	0.60***	0.58***	0.62***	0.51***	0.63***	0.65***
Text:TF-IDF	0.47***	0.45***	0.48***	0.38***	0.49***	0.52***
Text:TF-IDF-50	0.17***	0.14**	0.2***	0.18**	0.17**	0.08
Text:TF-IDF-L	0.47***	0.44***	0.49***	0.38***	0.49***	0.49***
Time:Days	0.22***	0.20***	0.24***	0.17**	0.25***	0.23***
Section:JACC	0.49***	0.47***	0.50***	0.36***	0.58***	0.62***
Tags:JACC	0.33***	0.36***	0.30***	0.25***	0.25***	0.45***
Title:BERTopic	0.30***	0.28***	0.32***	0.20***	0.24***	0.35***
Title:LDA	0.07*	0.04	0.10	0.04*	0.20***	0.05
Title:SBERT	0.38***	0.38***	0.39***	0.35***	0.45***	0.41***
Title:TF-IDF	0.20***	0.19***	0.2***	0.09	0.16**	0.28***
Title:TF-IDF-L	0.17***	0.15***	0.18***	0.09	0.11	0.20**
Title:BI	0.18***	0.19***	0.16***	0.16**	0.13**	0.21***
Title:JW	0.21***	0.2***	0.21***	0.14*	0.23***	0.26***
Title:LCS	0.22***	0.27***	0.17***	0.19**	0.22***	0.35***
Title:LV	0.18***	0.19***	0.16***	0.16**	0.12*	0.22***



T-test Results:

Group	All participants		National participants		Local participants	
	T	p	T	p	T	p
All	1.896	0.060	2.139	0.041	1.094	0.277
5	1.284	0.204	0.370	0.715	1.253	0.214
4	2.801	0.006	1.591	0.124	2.300	0.024
3	-1.313	0.118	0.418	0.681	-1.584	0.118
2	-1.682	0.480	-0.721	0.480	-1.554	0.124
1	0.575	0.566	0.926	0.363	0.0	1.0



Two news articles as they were shown in the survey

PARTNERS



HOST



FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

