User Satisfaction with Text-Based News Recommender Systems

Keywords: News Diversity, News Recommender Design, Serendipity, Similarity Based Recommendations, User Satisfaction

Extended Abstract

Journalistic media are increasingly responding to changes in user behaviour online by implementing algorithmic recommendations on their pages (Kunert & Thurman, 2019). However, unlike digital platforms, news providers have neither the ability nor the interest to mimic the data-hungry implicit personalization strategy of social media (Eskens, 2019;). Therefore, basing article recommendations on text similarity can be a suitable strategy for these organizations. From a societal viewpoint, recommendations should then be as diverse as possible . However, users tend to prefer recommendations that enable “serendipity” - the perception of an item as a welcome surprise that strikes just the right balance between more of similarly useful but still novel content. Having this possible trade-off in mind, the study explores (RQ1) how the satisfactory a text-based news recommendation algorithm is varying depending on the degree of text similarity, (RQ2) how satisfaction is related to evaluations of the recommended article as a) novel and b) unexpected and whether (RQ3) there are user dispositions that might impact this satisfaction, making it necessary to optimize the content-based news recommendation engine for specific user groups.

Research Design. The study implements an innovative research design, integrating a real recommendation engine for news articles into a survey; even though the advantage of combining web tracking and survey data is clear (Bernstein et al., 2020), mainly the big tech platforms have taken advantage of it so far (Stray, 2020). Starting with questions about news usage, an interactive part follows in which the respondents freely search a database of actual news articles. Participants enter a search query that is of interest to them and related to politics, business, or culture, then selecting one of the multiple search results for further reading. Immediately after reading, participants evaluate the quality of the self-selected article. Afterwards, another article is automatically recommended for further reading. Recommendations are based on the content of the previously selected texts aiming to present similar texts. Similarity is calculated based on vector representations of each article – which are in turn based on the average of the vectors of each sentence (sentence embeddings), calculated with the pre-trained language model BERT developed by Google (Devlin et al., 2018). Thus, quantification of a statistical similarity between all articles is possible and represented as a similarity score (values between 0, no similarity, and 1, identical). The three recommendation logics were randomly assigned to the participants. To generate a representative news corpus, the URLs of relevant texts from ten different media (e.g., Bild, Sueddeutsche Zeitung, F.A.Z) were first saved via News API (2021) and scraped in the next step. These 194,167 German news texts from the year 2020 were then vectorized. Participants represent all German-speaking Internet users (18+ yrs.) and were recruited via a commercial online access panel, cross-sampled according to education and age, place of residence, gender. After data cleaning, N = 588 participants remain (average age 48.2 yrs., 51.5 percent male, 53.9 percent with a low level of formal education), offering a good representation of the German online population.

Results. Using a series of regression analyses (see Table 1; block wise approach with forced entry), our findings show that the degree of text similarity between the original and the recommended article turns out to be the strongest predictor of the entire model. The stronger the recommendation is based on article similarity, the more pleasant and enriching it is perceived to be (RQ1). Moreover, this satisfaction strongly depends on whether the recommended article is evaluated as novel (RQ2a). By contrast, if the topic, facts, and/or viewpoints of the recommended article are perceived as unexpected, this decreases satisfaction with the recommendation (RQ2b). This confirms previous research on the preference of news users for "reliable surprises" (Schoenbach, 2007), news media should aim to recommend-and produce-news content that adds novel positions or facts, but still falls within the expectations of users for the topic. In all models, the more educated rate article recommendations as less pleasant and enriching (RQ3). Yet, higher news interest leads to slightly higher recommendation satisfaction. Users with a high Need for Cognitive Closure are also more likely to be satisfied, which stands to reason given that the recommendation is based on text similarity. For general attitudes toward news personalization, a significant simple effect emerges - endorsing algorithmic personalization leads to greater recommendation satisfaction.

Conclusion. Our results indicate that news organizations cannot go much wrong with a text-based recommendation algorithm, as indicated by the strength of the predictor text similarity on recommendation satisfaction. However, we find several individual dispositions for which this relationship is less strong – calling for a minimum of personalization in article recommendations. And if we not only focus on the "liberal-individualist" goal of satisfaction (Helberger et al., 2018), but at the same time consider the aspect of unexpected and thus potentially challenging content (which is important from a deliberative point of view), a trade-off becomes apparent. Here, the concept of serendipity in information system design, also known as “reliable surprise” in communication science, might offer a valuable link between a sometimes challenging diversity in news and yet pleasant user experiences and deserves further research.

References

Bernstein, A., Vreese, C. d., Helberger, N., Schulz, W., Zweig, K., Baden, C., . . . Zueger, T. (2020, May 19). Diversity in news recommendations. Retrieved from http://arxiv.org/pdf/2005.09495v1

Eskens, S. (2019). A right to reset your user profile and more: GDPR-rights for personalized news consumers. *International Data Privacy Law*, 9(3), 153–172. https://doi.org/10.1093/idpl/ipz007

Helberger, N., Karppinen, K., & D’Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society, 21*(2), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

Kunert, J., & Thurman, N. (2019). The form of content personalisation at mainstream, transatlantic news outlets: 2010–2016. *Journalism Practice, 13*(7), 759–780. https://doi.org/10.1080/17512786.2019.1567271

Schoenbach, K. (2007). ‘The own in the foreign’: reliable surprise - an important function of the media*? Media, Culture & Society, 29*(2), 344–353. https://doi.org/10.1177/0163443707074269

Stray, J. (2020). Aligning AI optimization to community well-being. *International Journal of Community Well-Being, 3*(4), 443–463. https://doi.org/10.1007/s42413-020-00086-3

Table 1: Hierarchical regression predicting recommendation satisfaction

	Model 1				Model 2				Model 3			
	b	*b* 95% CI [LL, UL]	*sr²*	*sr²* 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	*sr²*	*sr²* 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	*sr²*	*sr²* 95% CI [LL, UL]
Independent Variables												
Intercept	**2.71****	[2.57, 2.84]			**2.40****	[2.09, 2.72]			**2.61****	[2.28, 2.95]		
Block 1: Demographic variables												
Gender (1 = male)	**0.17***	[0.01, 0.34]	.01	[-.01, .02]	**0.19****	[0.05, 0.32]	.01	[-.00, .02]	**0.14***	[0.01, 0.28]	.00	[-.00, .01]
Age	**-0.01***	[-0.01, -0.00]	.01	[-.01, .02]	**-0.00**	[-0.01, 0.00]	.00	[-.00, .01]	**-0.00**	[-0.01, 0.00]	.00	[-.00, .01]
Education (1 = high)	**-0.14**	[-0.31, 0.02]	.00	[-.01, .02]	**-0.23****	[-0.36, -0.09]	.01	[-.00, .03]	**-0.24****	[-0.38, -0.11]	.01	[-.00, .03]
Block 2: Recommendation measures												
Similarity score					**0.70****	[0.28, 1.11]	.01	[-.00, .03]	**0.72****	[0.31, 1.13]	.01	[-.00, .03]
Quality article 1					**0.07***	[0.01, 0.13]	.01	[-.00, .02]	**0.03**	[-0.03, 0.09]	.00	[-.00, .01]
Novelty of recommendation					**0.67****	[0.58, 0.75]	.28	[.22, .34]	**0.63****	[0.55, 0.71]	.24	[.18, .29]
Unexpectedness of recommendation					**-0.22****	[-0.30, -0.14]	.03	[.01, .06]	**-0.20****	[-0.28, -0.13]	.03	[.01, .05]
Block 3: Individual dispositions												
Attitude towards personalization									**0.20****	[0.11, 0.29]	.02	[.00, .04]
Relative share of algorithmically personalized news									**-0.19**	[-0.49, 0.11]	.00	[-.00, .01]
Duty to keep informed (DTKI)									**-0.06**	[-0.21, 0.08]	.00	[-.00, .00]
Need for Cognitive Closure (NfCC)									**0.12***	[0.02, 0.22]	.01	[-.00, .02]
News interest									**0.14***	[0.02, 0.26]	.01	[-.00, .02]
News trust									**0.01**	[-0.07, 0.10]	.00	[-.00, .00]
Technical affinity									**-0.03**	[-0.11, 0.06]	.00	[-.00, .00]
R2	**R2 = .016***				**R2 = .351****				**R2 = .387****			
95% CI	[.00,.04]				[.28,.40]				[.31,.43]			
ΔR2					**ΔR2 = .335****				**ΔR2 = .036****			
95% CI					[.27, .40]				[.01, .06]			

*Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05. ** indicates p < .01.*