Modeling the Dynamics of Online News Reading Interests

Elena Viorica Epure, Benjamin Kille, Jon Espen Ingvaldsen, Rebecca Deneckere, Camille Salinesi, Sahin Albayrak

▶ **To cite this version:**

HAL Id: hal-01516160
https://paris1.hal.science/hal-01516160v2

Preprint submitted on 6 Feb 2023

Modeling the Dynamics of Online News Reading Interests

Elena Viorica Epure
CRI, Panthéon-Sorbonne
Paris, France
Elena.Epure@malix.univ-paris1.fr

Benjamin Kille
DAI-Lab, TU Berlin
Berlin, Germany
benjamin.kille@dai-labor.de

Jon Espen Ingvaldsen
IDI, NTNU
Trondheim, Norway
jonespi@idi.ntnu.no

Rebecca Deneckere
CRI, Panthéon-Sorbonne
Paris, France
Rebecca.Deneckere@univ-paris1.fr

Camille Salinesi
CRI, Panthéon-Sorbonne
Paris, France
Camille.Salinesi@univ-paris1.fr

Sahin Albayrak
DAI-Lab, TU Berlin
Berlin, Germany
sahin.albayrak@dai-labor.de

ABSTRACT

Online news readers exhibit a very dynamic behavior. News publishers have been investigating ways to predict such changes in order to adjust their recommendation strategies and better engage the readers. Existing research focuses on analyzing the evolution of reading interests associated with news categories. Compared to these, we study also how relations among news interests change in time. Observations over a 10-month period on a German news publisher indicate that overall, the relations amid news categories change, but stable periods spanning months are also found. The reasons of these changes and how news publishers could integrate this knowledge in their solutions are subject to further investigation.

KEYWORDS

news reading dynamics; temporal analysis; news reading interests

1 INTRODUCTION

Publishers employ recommender systems to target online content and engage readers. They track reading interests to continuously infer strategies to match readers and content. Further, as events emerge, journalists add news articles to publishers' repertoires. This creates a highly dynamic environment where future interactions cannot be reliably predicted. In this work, we analyze whether relations among news categories exhibit regularities as they evolve. Such regularities, though they exist, would represent valuable information which publishers could exploit as competitive advantage. We reduce the complex problem to two rather simple questions: Which news category will users follow up reading articles of a given category? Will these transitioning patterns remain stable over time? This lets us devise a probabilistic approach to model the evolution of reading behavior. Our contribution is novel as it explicitly considers transitions amid categories.

2 RELATED WORK

In the news domain, online readers rarely express their reading interests explicitly. Nonetheless, recommender systems often rely on such knowledge. In absence of this kind of information, publishers infer interests for news categories. In some cases, news category taxonomies are created in advance. Then, representative categories are selected for each article and included in the article's meta-data [3, 5]. In other cases, news categories are automatically identified from the article's content as either granular vocabularies [1] or broad news topics such as tourism or local news [6, 7].

Various works have analyzed the evolution of readers' interests from logs [6, 7] to improve news recommenders. Liu *et al.* [7] analyzed the Google News service. For each reader, the monthly average number of clicks per category was calculated. Then, the differences between the click distribution of the last assessed month and each of the previous months were plotted. They found that affinity to categories changed over time. Further, Li *et al.* [6] segmented the click history of news readers in consecutive time-frames from which they extracted the long- and short-term interests. The news interests of consecutive time-frames were compared through the KL-divergence score for each reader and the crowd average score plotted. Their evaluation indicates that short-term interests fluctuate, whereas, the long-term ones remained relatively stable.

Compared to these works, we drive the creation of news recommenders by assessing the progress of relations amid news interests. Specifically, we build on the analysis of Esiyok *et al.* [4] and we focus specifically on the temporal stability of category transitions.

3 EXPERIMENTS

Let $\mathcal{R} = \{r_m\}_{m=1}^{M}$, $\mathcal{A} = \{a_n\}_{n=1}^{N}$ refer to a news publisher's sets of readers and articles. The publisher observes the reading activity over time. Let $e = (e_r, e_a, e_t)$ define a reading event where each variable refers to reader, article, and time. For each reader, the publisher determines a sequence of events: $\Theta_r = \{e^{(k)} : e_r = r\}_{k=1}^{K}$ such that $e_t^{(k)} < e_t^{(k+1)}, \forall k \in K$. Let C refer to the set of news categories ($|C| = V$) and $c(e_a)$ to the category assigned to the event's article. The sequence of events associated with a specific reader r can be then transformed into a sequence of categories $C_u = (c(e_a^{(1)}), c(e_a^{(2)}), \ldots, c(e_a^{(K)}))$. If we consider C_u generated by a discrete first-order Markov Chain over news categories then the transition probabilities are represented by the matrix:

$$S = \begin{bmatrix} p_{11} & \cdots & p_{1V} \\ \vdots & \ddots & \vdots \\ p_{V1} & \cdots & p_{VV} \end{bmatrix} \quad p_{vw} \geq 0, \sum_{c_w \in C} p_{vw} = 1, \quad \forall v, c_v \in C.$$

We obtain p_{vw} as we divide the number of transitions from c_v to c_w by the number of transitions from c_v to any category.

Consequently, we can infer the evolution of reading behavior by observing the stationarity of S. Let t refer to an individual month, and T to the entire period. Let S_t refer to the transition matrix based on observations in $t \in T$. Let S_T refer to the transition matrix based on all observations. We apply a difference test based on the χ^2 statistic comparing each S_t with S_T [2]. If the test yields non-stationarity for the entire period then the computed values per month are plotted for getting insights into possible similar sub-periods. Then the test is applied again for those.

The evaluation uses event logs from January–October, 2015, from a German news publisher with 3 million weekly visits on average. Editors manually assign categories to articles. Events pass through a pre-processing pipeline: determining session reference, sorting chronologically, and splitting in reading episodes. Splitting connects events if they occurred within 1 min to 60 min apart (average of 20 min per reading is reported[1]). Finally, we computed each S_t and the S_T. If articles of successive events had multiple categories assigned, each element of the Cartesian product of categories counts as a transition. The χ^2 test is sensitive to even minor variations which we expect given the large data set. We multiplied the frequencies by a factor of 10^{-3} to counteract the sensitivity.

The results show that the reading behavior for the entire period is not stable ($Q = 5994.6, \mathrm{df} = 889, p = 0$). However, stationary sub-periods were identified: January–Mars ($Q = 301.65, p = 0.038, \mathrm{df} = 260$); April–July ($Q = 6.56, p > 0.99, \mathrm{df} = 48$); September–October ($Q = 14.95, p > 0.99, \mathrm{df} = 55$), and August is a singularity. Figure 1 illustrates transition probabilities for three news categories over the 10-month period. Point sizes reflect the probability. For instance, we observe that transitions remain mostly stable over the period April–July 2015. This yields for all three categories visible through the horizontally aligned points. At the same time, we observe times when transitions change significantly. Comparing the period January–March with April–July, we find hardly any similarity.

We conclude that there are rather stable periods whose connections exhibit irregular breaks in transition probabilities. As a result, preceding months appear to be an adequate basis to define strategies predicting future reading behaviors unless the reading patterns change drastically. We contemplate that various factors could cause such abrupt changes. These may include readers' changing interests due to seasons, publishers adjusting their websites, and breaking news events with long-term effects. Recognizing these fluctuations early on represents a major challenge publishers will have to address in the future. The ability to adjust quickly to major changes could become a decisive competitive advantage for publishers.

4 CONCLUSION AND PERSPECTIVES

Reading interests change over time. Still, we observed close to stationary transition probabilities for periods stretching multiple months. This indicates that publishers can exploit reading behavior to improve targeting content to readers and increasing activity on their platforms. We expect other factors to affect how well the system matches readers and contents. We observed breaks in reading behavior on multiple occasions. Anticipating these inconsistencies
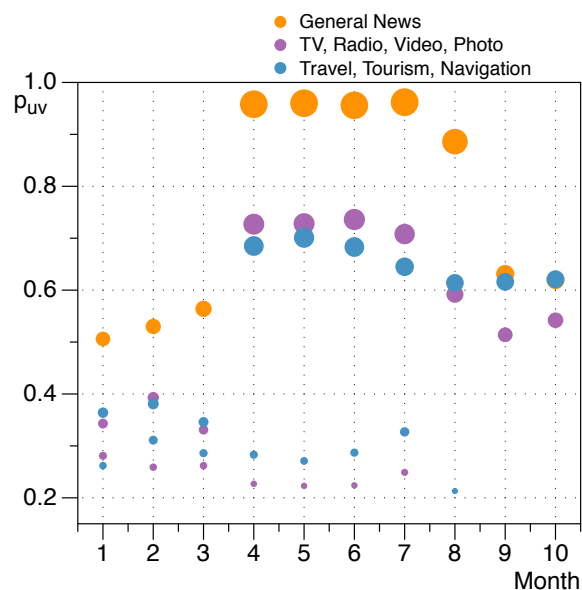

Figure 1: Transition probabilities per month from *General news* (orange), *Travel, Tourism, Navigation* (blue), *TV, Radio, Video, Photo* (violet) to the other categories between January and October 2015. Probabilities > 0.2 are displayed only.

creates another issues publishers will have to address. Further analysis is necessary to develop a broader understanding for the degree to which the categories determine reading behavior.

REFERENCES

[1] Amr Ahmed, Choon Hui Teo, S.V.N. Vishwanathan, and Alex Smola. 2012. Fair and Balanced: Learning to Present News Stories. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining.* ACM, New York, NY, USA, 333–342. DOI:https://doi.org/10.1145/2124295.2124337

[2] F. Bickenbach and E. Bode. 2003. Evaluating the Markov Property in Studies of Economic Convergence. *International Regional Science Review* 26, 3 (2003), 363–392. DOI:https://doi.org/10.1177/0160017603253789

[3] E. V. Epure, J. E. Ingvaldsen, R. Deneckere, and C. Salinesi. 2016. Process mining for recommender strategies support in news media. In *Proceedings of the 10th International Conference on Research Challenges in Information Science.* 1–12. DOI:https://doi.org/10.1109/RCIS.2016.7549356

[4] Cagdas Esiyok, Benjamin Kille, Brijnesh-Johannes Jain, Frank Hopfgartner, and Sahin Albayrak. 2014. Users' Reading Habits in Online News Portals. In *Proceedings of the 5th Information Interaction in Context Symposium.* ACM, New York, NY, USA, 263–266. DOI:https://doi.org/10.1145/2637002.2637038

[5] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. 2010. A Contextual-bandit Approach to Personalized News Article Recommendation. In *Proceedings of the 19th International Conference on World Wide Web.* ACM, New York, NY, USA, 661–670. DOI:https://doi.org/10.1145/1772690.1772758

[6] Lei Li, Li Zheng, Fan Yang, and Tao Li. 2014. Modeling and Broadening Temporal User Interest in Personalized News Recommendation. *Expert Systems with Applications* 41, 7 (2014), 3168–3177.

[7] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized News Recommendation Based on Click Behavior. In *Proceedings of the International Conference on Intelligent User Interfaces.* 31–40.