Interactive Topic Modeling for the Broadcasting Media

Laura Ham*
Independent researcher
The Netherlands

Luis A. Leiva
University of Luxembourg
Luxembourg

ABSTRACT

Broadcasting companies produce large amounts of text and audiovisual content. Extracting meaningful insights from these sources requires efficient analysis methods, which are often only palatable to data scientists. Even in large organizations there is a critical *knowledge gap*: media experts manually curate work to derive insights, which is very time consuming, while engineers can use advanced data science methods but lack the domain expertise to derive key insights from the data. We propose to bridge this knowledge gap with INTEX, a human-in-the-loop interactive topic modeling application. We designed INTEX considering non-technical media experts as the main stakeholders of the application. A user evaluation shows that INTEX enables domain experts to extract and explore topics in an intuitive and efficient manner. Our work illustrates how complex applications can be made more accessible by hiding low-level details and linking these to high-level interpretations.

CCS CONCEPTS

• **Computing methodologies** → *Topic modeling*; • **Human-centered computing** → *Interactive systems and tools*; **User centered design**.

KEYWORDS

Interactive Machine Learning; Human-in-the-loop; Topic Modeling; Exploratory Data Analysis

ACM Reference Format:

Laura Ham and Luis A. Leiva. 2023. Interactive Topic Modeling for the Broadcasting Media. In *XXIII International Conference on Human Computer Interaction (Interaccion 2023), September 4–6, 2023, Lleida, Spain.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3612783.3612803

1 INTRODUCTION

Broadcasting media companies produce large amounts of text and audiovisual content worth of analysis. For example, the Washington Post produces about 500 stories per day [20] and Netflix has 2.2 million minutes of content, or over 50,000 titles, only in the US [21]. Tapping these sources helps to uncover hidden patterns and gain insights to support data-driven business decisions. This requires efficient analysis methods and modeling techniques for automatic theme discovery, among which *topic modeling* is the most popular one. In a nutshell, topic modeling infers latent structures of large document collections by automatically coding them into a smaller number of semantically meaningful categories.

A shortcoming of classic topic models is that the discovered topics can be hard to interpret [12]. Likewise, extracting too many or too few topics leads to either too general or too specific results [12]. To solve these issues, Interactive Topic Modeling (ITM) has been introduced recently, which incorporates human expertise in the modeling process [14]. ITM applications allow users to refine extracted topics by e.g. keyword and document source. These applications are typically used by data scientists, who are experienced in Natural Language Processing (NLP). However, these NLP experts often lack domain knowledge about the data and its high-level interpretation in a business context. At the same time, domain experts in the broadcasting media, like journalists and data analysts, have this broader knowledge about the produced and consumed media content, but usually lack data science skills to develop and use complex topic models.

This knowledge gap between data scientists and domain experts is excruciating, because strategies for thinking and problem solving differ significantly [26] and also because domain experts find it hard to articulate their problems [25]. We propose to fill this knowledge gap with INTEX (INteractive Topic EXplorer), a human-in-the-loop ITM application designed according to Human Centered Design (HCD) principles [11] in collaboration with end-users.

2 RELATED WORK

Topic models typically reduce the dimensionality of a set of words in a set of documents into a smaller set of interpretable and meaningful themes, or topics. Documents many cover several different topics whereas words can be associated with multiple topics. Classic approaches to topic modeling include Latent Semantic Analysis (LSA) [9] and variations thereof such as pLSA [13]. While these approaches may create compact semantic representations [29], they are not attractive for real-world use cases because the discovered topics and keywords are hard to interpret. More recently, Latent Dirichlet Allocation (LDA) was shown to discover more descriptive topics [5], however LDA is suboptimal in terms of consistency and convergence [8]. Both model consistency and convergence are important from the user's point of view, as low consistency and slow model convergence lead to bad user experience. Non-negative Matrix Factorization (NMF) overcomes these aforementioned problems [8], leading to outcomes that are naturally interpretable [2, 23] in a computationally efficient way [29], so it is preferred over other topic models in practice.

An intuitive representation of the extracted topics, as well as the underlying model, is desired to promote understanding, since in ITM applications the user can control modeling results by direct manipulation. Pevious work presented results as word lists [7, 30], word clouds [10], bubble charts [22], and Sankey diagrams [27].

*Work done while affiliated with Aalto University.

Interaccion 2023, September 4–6, 2023, Lleida, Spain

ACM ISBN 979-8-4007-0790-2/23/09.
https://doi.org/10.1145/3612783.3612803

Lee et al. [17] concluded that topics may be misinterpreted because of the words representing them, and recommended that topic refinement should be focused on topics with low coherence [6].

Effective collaboration in ITM requires both transparency and predictability [1, 16]. However, there is often a trade-off observed between the two, since high transparency, where model outcomes are easy to validate, expects predictable outcomes and makes it difficult to provide users with suitable controls [28]. Therefore, ITM applications need to balance user controls and truly modeling the data, in order to promote trust in the application [3]. To the best of our knowledge, there is no ITM application for broadcasting media companies. Further, since domain and user expertise largely impact how a topic model is perceived and used, they should be considered in the design of any ITM application.

3 SYSTEM

Following previous work and the HCD principles, a formative user study with stakeholders was conducted to gather business requirements. Later on, participatory design methods were applied, where the users interacted with a product prototype for evaluation purposes. Then, recurring one-hour sessions with focus groups were organised, using a combination of three user study techniques: *Concept testing*, *Desirability studies*, and *Participatory design*. The three techniques are attitudinal, mostly qualitative, and all incorporate hybrid prototype usage during data collection [24].

3.1 Design Choices and Interactions

INTEX's interface is designed according to the mental model of a non-technical end-user, which is presented in Figure 1. The workflow of the application is summarized as follows.

(i) **Data input selection**. Users can select sources, filter by metadata, and choose a time window. Immediate feedback to user's input is presented either in graphical or tabular form.

(ii) **Model configuration**. The only hyperparameter in INTEX is the number of topics to extract. Since choosing the optimal number of topics beforehand can be challenging, a topic suggestion of 15 topics is provided initially.

(iii) **Model interpretation and assessment**. Model output is shown as topic-term and document-topic tables, to provide a quick overview of the generated topics. Users can notice what topics are reflected in the input set of documents and can see suggestions on what topics need refinement.

(iv) **Model refinement**. Iterations with focus group sessions resulted the following set of options: merge, split, remove keywords from a topic, and rename a topic. INTEX's visualizations reflect the results of these refinements in real-time.

(v) **Exploratory data analysis (EDA) in wider context**. Users can see topic development over time and compare topic content production with consumption by different age groups. Data export, including intermediate modeling steps, is also available.

As noted, the workflow in INTEX covers the whole 'user journey', from data selection to exploratory data analysis and exporting the results. The user can follow the five steps explained above both in sequential order or can go back to any earlier step at their own will, for example to change the data input or model configuration after model refinement.

3.2 Implementation

INTEX is delivered as a web-based single-page application. Fast and smooth transitions between pages make the web application feel like a native desktop-based app. The back-end of INTEX uses the Python library Stanza[1] on top of SpaCy[2] to preprocess text documents and apply NMF for topic modeling. The front-end of INTEX is built with the open-source framework *Streamlit*.[3] The user interface has two main modules (Figure 1a). A panel on the right displays information in the form of text, tables, and interactive visualizations. A panel on the left takes in the user actions that influence the topic model.

Regarding model configuration, INTEX includes a bar chart of the 15 most prevalent keywords in the document corpus, as well as a table showing the document frequency of each word. The user can also select the number of topics to extract and click on a button to run initalize the model (Figure 1b). As soon as the topics are derived, the interface allows the user to interpret the model, refine it, explore the resulting data, and export them. The interactive topic visualization to explore relationships between topics and keywords is only generated as per-user request. This visualization is made using LDAvis,[4] which generates an interactive HTML file from the output of a topic model. Relations between individual documents and their topics are visualized with the Bokeh library[5] and the UMAP dimensionality reduction method [19].

INTEX allows the user to get an estimation of the topics quality, by comparing model residuals per topic, as well as refine topics on demand. Note that topics are usually independent from geographical location. Finally, exploratory data analysis and data export options are also available (Figure 1c).

4 EVALUATION

We analyzed the perceived usability (efficiency and satisfaction) and user experience of INTEX via rating scales that were complemented with semi-structured interviews at post-task. A dataset of 605 articles were provided by Yle, the Finnish national broadcasting company.[6] Each article comprises $M = 318$ words ($SD = 231$) after text preprocessing. Ten participants (6 female, 4 male) aged 30–39 were recruited from Yle. They have various backgrounds regarding data analytics and data science.

We conducted individual evaluation sessions that took up to one hour per participant. Each session was conducted remotely with audio and screen-capture recording. Each session started with a walk-through of INTEX. Then, the following task scenario was presented: *"You want to make a report about the most important news articles published by Yle's Current Affairs department in the last year. Use INTEX to derive a set of topics that would help you and your target audience to understand the contents that have been covered by such news articles."* Participants were instructed to think-aloud during this task. Afterwards, a short semi-structured interview was conducted.

[1] https://stanfordnlp.github.io/stanza/
[2] https://spacy.io/
[3] https://www.streamlit.io/
[4] https://github.com/bmabey/pyLDAvis
[5] https://bokeh.org/
[6] https://yle.fi/

(a) Data input and filtering (b) Model output (c) Exploratory visualizations

Figure 1: Screenshots of the main sections of INTEX. Here, the journalist is analyzing the impact of a set of international news during 2019.

We logged task completion time along with the aforementioned post-test survey. Perceived usability was measured in a 1–5 Likert scale (1: 'strongly disagree', . . . , 5: 'strongly agree') following the System Usability Scale (SUS) [15]. User experience was measured with nine questions adopted from Smith et al. [2020]: The first four questions relate to frustration, trust, task ease, and confidence. The next five questions relate to model adherence, instability, latency, quality, and improvement. Participants answered these questions again on a 1–5 point Likert scale. Finally, the outcomes of the think-aloud protocol were coded thematically.

4.1 Results

Regarding task completion time, participants spent $M = 24$ minutes ($SD = 7$ minutes) on the task. This represents a significant improvement of over 50% with regards to manual work, which all participants estimated to be more than one hour at the very least. We should point out that none of the participants had conducted any topic modeling task before, but quickly understood and saw the value of INTEX for their daily work.

Regarding system usability and user experience, the average SUS score is 81, which is well above the benchmarked average for websites and web applications [4, 18]. Participants found the task easy ($M = 4.2$, $SD = 0.8$), trusted INTEX ($M = 4.1$, $SD = 0.7$), felt confident using it ($M = 3.9$, $SD = 0.9$), and did not experience frustration (negative statement, lower is better, $M = 1.8$, $SD = 0.8$). Participants mentioned that INTEX adhered to their input ($M = 4.2$, $SD = 0.9$), had low latency ($M = 4.1$, $SD = 0.9$), and was not unstable (negative statement, $M = 1.8$, $SD = 0.6$). In addition, participants argued that the final topics were substantially improved over the initial topics ($M = 3.8$, $SD = 0.4$) and most participants were satisfied with the results ($M = 4.2$, $SD = 0.4$).

4.2 Research Findings

We distill the most relevant findings from the post-task interviews.

(1) **Users do not change the model configuration before an initial run.** Only two participants made changes to the default model configuration before they ran the model for the first time. Three participants made no changes at all.
(2) **Users are uncertain about the amount of topics to extract, but only initially.** Three participants decided to run the model with the default number of 5 and four participants specified from 10 to 20 topics.
(3) **Users like to explore and try out functionality multiple times.** Seven participants ran the model with at least two different numbers of topics to extract. This suggests that users engaged with INTEX instead of just following the study instructions blindly.
(4) **Exploratory data visualizations do not only serve as an 'extra' analysis, but also support the users in topic model understanding, evaluation, and refinement.** Four participants have used the visualizations to evaluate individual topic results, and made model refinements based on analysing these graphs.
(5) **Users like the variety and interactivity of the visualizations.** All the participants pointed out that they liked the fact that the visualizations are interactive. For example, P6 mentioned: *"When I hover over data in the charts, I see more information on the content and the topic clusters, that is really good."*.
(6) **INTEX triggers users' curiosity and encourages them to use the application with other data sources.** When exploring the last plot including consumption data per topic, P5 stated that *"I want to do some analysis on questions that come to my mind. It is great that I can do this analysis with just a few clicks."*
(7) **Users like to be in control of the model, but would like to see suggestions and get a sense of its quality.** Users got excited when they see that manual refinements to the model are implemented and reflected as expected. Users also appreciate recommendations made by the application about which topics to refine.
(8) **Users like to see model refinements reflected immediately.** Multiple participants verbally appreciated the results of user and

system actions to be directly visible. P3 indicated that *"the value of INTEX is that you get really quickly information on the data"*.

(9) **Users without previous modeling knowledge feel confident using INTEX.** Participants indicated that they are amazed by how much they can achieve with this application without having any technical experience. They did not feel the need to completely understand the underlying model, but indicated their confidence in their actions and the final results.

(10) **Users like to explore functionalities rather than reading any user manual first.** None of the participants conducted the 'how to use' tutorial before they started the modeling task. P10 said: *"I know what the application should be roughly capable of, so I will just explore the functionalities"*.

5 DISCUSSION

Our results indicate that INTEX indeed helps non-technical end-users perform topic modeling and intervene on the process, without the help of data scientists. A SUS score of 81 indicates that the applications has a high usability, falling in the top 10% of scores (90th percentile), suggesting that INTEX has 'excellent' performance [18] in terms of effectiveness, efficiency, overall ease of use, and learnability. For example, participants indicated that being able to control the model results by the offered refinement options does not only influence their user experience but also has a positive impact on how they perceived the final results.

We found that participants demand different types of visualizations, depending on their background and personal interests. None of the participants indicated that there were too many visualizations, or that visualizations were too complicated. Although our participants had mixed expertise, the user experience and perception results are quite stable. The average standard deviation over all questions is 0.71 on a five-point Likert scale, with the highest standard deviation of 0.9 for statements on user confidence, perceived model adherence, and model latency.

Limitations. INTEX is unique in its combination of use context, data, users, interface design, topic model and visualization techniques. Currently there are no existing ITM applications for the broadcasting domain, and previous work has not designed with and for domain experts, so we cannot compare INTEX against any competing ITM application. In addition, previous ITM applications have not been evaluated with real users under a similar setting like ours, so it is difficult to compare results across domains. This implies that the results of our user evaluation are indicators of the usability and user experience with INTEX only.

6 CONCLUSION

INTEX is an interactive topic modeling application for media content production analysis that bridges the gap between domain experts (who lack data science knowledge) and data scientists (who lack expert domain knowledge). Based on a formal user evauation, we can conclude that INTEX is highly usable, promotes an adequate user experience, and is easy to learn by non-technical domain experts. Our software is publicly available at https://github.com/laura-ham/INTEX.

ACKNOWLEDGMENTS

We thank Yle for their support. Work supported by the Horizon 2020 FET program of the European Union (grant CHIST-ERA-20-BCI-001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

REFERENCES

[1] B. Abdollahi and O. Nasraoui. 2018. Transparency in fair machine learning: The case of explainable recommender systems. In *Human and Machine Learning*.
[2] S. Arora, R. Ge, and A. Moitra. 2012. Learning topic models–going beyond SVD. In *Proc. FOCS*.
[3] A. Bakharia, P. Bruza, J. Watters, B. Narayan, and L. Sitbon. 2016. Interactive topic modeling for aiding qualitative content analysis. In *Proc. CHIIR*.
[4] A. Bangor, P. T. Kortum, and J. T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact.* 24, 6 (2008).
[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, Jan (2003).
[6] G. Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proc. GSCL* (2009).
[7] A. J.-B. Chaney and D. M. Blei. 2012. Visualizing topic models. In *Proc. AAAI*.
[8] J. Choo, C. Lee, C. K. Reddy, and H. Park. 2013. Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. *IEEE Trans. Vis. Comput. Graph.* 19, 12 (2013).
[9] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. 1990. Indexing by latent semantic analysis. *J. Assoc. Inf. Sci. Technol.* 41, 6 (1990).
[10] M. J. Gardner, J. Lutes, J. Lund, J. Hansen, D. Walker, E. Ringger, and K. Seppi. 2010. The topic browser: An interactive tool for browsing topic models. In *Proc. NeurIPS Workshops*, Vol. 2.
[11] K. Goodwin. 2009. *Designing for the Digital Age: How to Create Human-Centered Products and Services.* Wiley Publishing, Inc.
[12] D. Greene, D. O'Callaghan, and P. Cunningham. 2014. How many topics? stability analysis for topic models. In *Proc. ECML-PKDD*.
[13] T. Hofmann. 1999. Probabilistic Latent Semantic Analysis. In *Proc. UAI*.
[14] Y. Hu, J. Boyd-Graber, B. Satinoff, and A. Smith. 2014. Interactive topic modeling. *Mach. Learn.* 95, 3 (2014).
[15] P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester. 1996. *Usability evaluation in industry.* CRC Press.
[16] T. Kulesza, S. Stumpf, M. Burnett, W.-K. Wong, Y. Riche, T. Moore, I. Oberst, A. Shinsel, and K. McIntosh. 2010. Explanatory debugging: Supporting end-user debugging of machine-learned programs. In *Proc. IEEE VL/HCC*.
[17] T. Y. Lee, A. Smith, K. Seppi, N. Elmqvist, J. Boyd-Graber, and L. Findlater. 2017. The human touch: How non-expert users perceive, interpret, and fix topic models. *Int. J. Hum. Comput. Stud.* 105 (2017).
[18] J. R. Lewis and J. Sauro. 2018. Item benchmarks for the system usability scale. *J. Usability Stud.* 13, 3 (2018).
[19] L. McInnes, J. Healy, and J. Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426* (2018).
[20] R. Meyer. 2016. How Many Stories Do Newspapers Publish Per Day? Available at theatlantic.com. Last accessed on Jan 2022.
[21] K. Moore. 2020. How Long Would It Take To Watch All Of Netflix? Available at whats-on-netflix.com. Last accessed on Jan 2022.
[22] J. Murdock and C. Allen. 2015. Visualization techniques for topic model checking. In *Proc. AAAI*.
[23] P. Paatero and U. Tapper. 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 2 (1994).
[24] C. Rohrer. 2014. When to use which user-experience research methods.
[25] M. Sedlmair, M. Meyer, and T. Munzner. 2012. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graph.* 18, 12 (2012).
[26] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. 2015. Bridging the gap of domain and visualization experts with a liaison. In *Proc. EuroVis*, Vol. 2015.
[27] A. Smith, J. Chuang, Y. Hu, J. Boyd-Graber, and L. Findlater. 2014. Concurrent visualization of relationships between words and topics in topic models. In *Proc. ACL Workshops*.
[28] A. Smith-Renner, V. Kumar, J. Boyd-Graber, K. Seppi, and L. Findlater. 2020. Digging into user control: perceptions of adherence and instability in transparent models. In *Proc. IUI*.
[29] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler. 2012. Exploring topic coherence over many models and many topics. In *Proc. EMNLP*.
[30] C. Zou and D. Hou. 2014. LDA analyzer: A tool for exploring topic models. In *Proc. ICSME*.