TOPIC TASSEOGRAPHER: AN INTERACTIVE R-SHINY APPLICATION TO INTERPRET AND VALIDATE TOPIC MODELS

Federico Mangiò – University of Bergamo, Department of Management;

Address: Via dei Caniana, 2, 24127 Bergamo (BG), Italy.

E-mail: Federico.mangio@unibg.it

ORCID: https://orcid.org/0000-0003-3997-4956

Preferred citation: Mangiò, F. (2026). Topic tasseographeR: An interactive R-Shiny application to interpret and validate topic models. https://doi.org/10.5281/zenodo.18815035

Abstract

Interpreting and validating topic-model outputs remains a persistent challenge in text-as-data research. We introduce Topic tasseographeR, an R Shiny application that supports topic interpretation and validation through post-hoc dictionary-based scoring. The app computes topic content and topic function scores by combining established lexical resources with topic–word and document–topic distributions. Designed to be model-agnostic, Topic tasseographeR integrates seamlessly with existing topic-modelling workflows and enables both deductive validation and abductive topic labelling. By leveraging validate dictionary-based methods, the software provides a complementary alternative to human- and machine-in-the-loop topic interpretation and validation approaches.

Keywords

R-shiny app; topic modelling; validation; sentiment analysis; text-as-data.

1. Introduction

In the wake of the linguistic turn [1], social scientists across different disciplines—including management, psychology, marketing, sociology, and political science, among others—have increasingly adopted topic models to analyse large text corpora [1,2]. Topic modelling is a bottom-up, automated approach to text analysis that comprises a family of algorithms designed to partition corpora into substantively meaningful clusters— “topics”—with minimal human intervention [3,4].

Topic models offer several advantages over traditional approaches to content analysis [4,5]. Compared to qualitative methods, they enable the systematic analysis of large textual corpora at scale. Compared to deductive approaches such as dictionaries and rule-based methods, they are inductive, as they do not require predefined categories. Moreover, they allow researchers to capture the relationality of meaning, by tracing how the sense of terms varies across contexts [1,5]. Overtime these advantages spurred a long and rapidly-evolving series of topic-modelling methods, which have evolved from generative probabilistic models based on bag-of-words representations, such as Latent Dirichlet Allocation (LDA; [6]), to more recent approaches that leverage embeddings, such as BERTopic [7].

Despite these methodological advances, a persistent challenge remains for researchers employing topic models: the interpretation and validation of topic-model outputs [8,9]. This task is often so difficult that it has been likened to “tasseography”, the art of reading tea leaves [10]. To address this challenge, various labelling and validation human-in-the-loop and machine-in-the-loop strategies have been proposed [9]. Human-in-the-loop approaches rely on post-measurement validation through human-computer comparisons, typically requiring human coders to manually annotate documents associated with each topic and compare their judgments to the model’s output [1,5]. While informative, these procedures are time-consuming, thereby undermining one of the core advantages of topic models, namely scalability, and often suffer from limited reliability. Machine-in-the-loop approaches, by contrast, incorporate semantic information or non-textual structure a priori into the topic-modelling process (e.g., seeded LDA [11] or Structural Topic Models [12]). Although useful, these approaches assume that seed dictionaries or theoretically relevant metadata are known in advance, an assumption that does not always hold in exploratory topic-modelling research. Beyond that, more recent proposals suggest replacing human coders with large language models (LLMs; [13, 14]). While promising, these approaches risk downplaying the role of the researcher and raise additional concerns related to cost, opacity, and reproducibility [15]. In this vein, the rapid pace of innovation in natural language processing risks encouraging researchers to prioritize technical novelty over validity concerns, leading to a quick dismissal of a long tradition of carefully validated lexical resources just to switch and adopt the most technically advanced tools [16].

To address these limitations, we introduce Topic tasseographeR, an R Shiny application designed to support the interpretation and validation of topic-model outputs through the computation and visualization of topic-level dictionary scores. Specifically, the app computes topic content and topic function scores for constructs measured by established dictionaries and lexica, accounting for both content and function words [17]. The approach implemented in Topic tasseographeR has already been used to interpret and validate topic models across multiple empirical contexts and social science domains [18,19,20,21], but has never been embedded in a user-friendly, ready to use interface. Overall, topic tasseographeR provides a fast, transparent, and reliable way to assist topic interpretation and validation, leveraging the long-standing tradition of dictionary-based analysis, and offering a complementary solution to existing human- and machine-in-the-loop validation strategies.

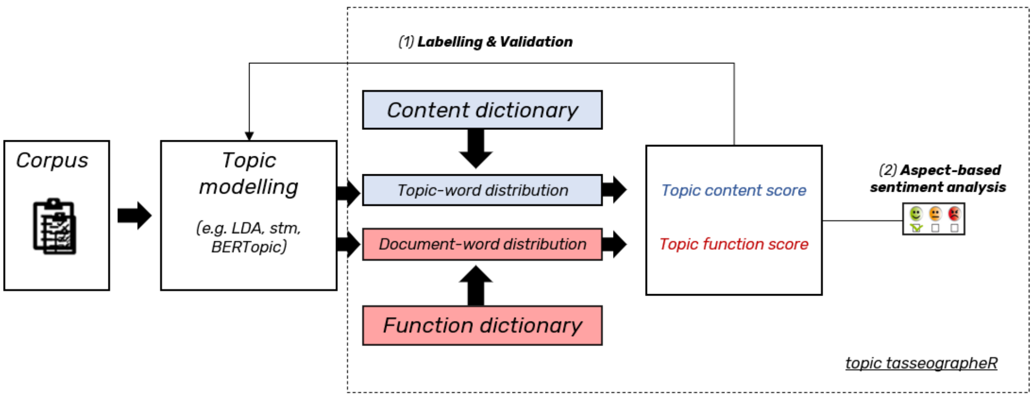
2. Method

2.1 Architecture

Topic tasseographeR is an interactive R Shiny application designed to assist researchers during topic-model interpretation and validation tasks. The software is deliberately model-agnostic and operates independently of any specific topic-modelling implementation. Its required inputs—namely the topic–word distribution and the document–topic distribution—can be generated using a wide range of topic-modelling tools implemented in R, Python, or visual analytics environments (e.g., KNIME Analytics).

The application consists of two complementary and symmetrical modules: "Topic *Content* TasseographeR" [shinyapps.io/topic_content_score/] and "Topic *Function* TasseographeR" [shinyapps.io/topic_function_score/]. Each module is dedicated to computing and visualizing topic-level dictionary scores based on the topic–word distribution and the document–topic distribution, respectively (Figure 1). Both modules are deployed as web applications and follow an identical user-interaction logic.

Figure 1 – Topic TasseographeR's protocol.

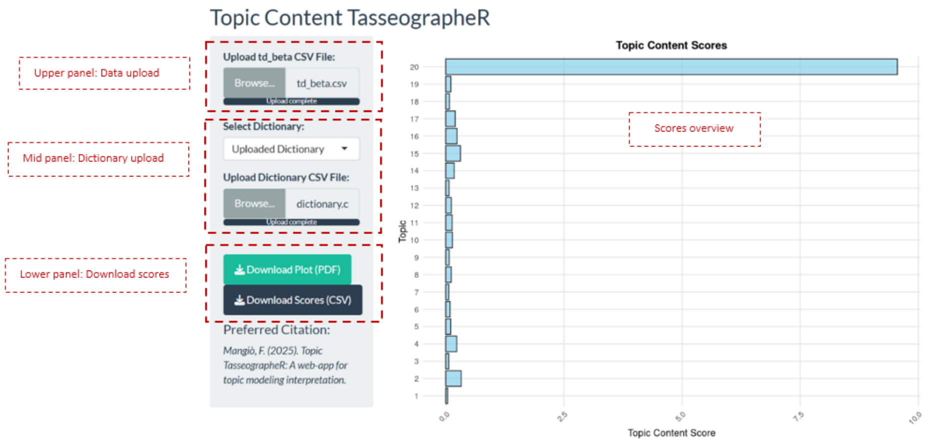


Data loading. In each module, users upload two locally stored comma-separated (csv) files via the top-left interface panel (Figure 2). The first file contains the topic–word distribution (β), referred to as the "td_beta" file, and must include three columns: (i) "topic", indicating the topic index, k; (ii) "term", containing the vocabulary terms, w; and (iii) "beta", representing the probability or word score $\beta_{k,w}$ of term w being associated with topic k. The second file contains the document–topic distribution (θ), referred to as the "ds" file, formatted as a matrix with one column containing the original document text and subsequent columns (Topic_k) reporting the estimated topic proportions θ_k for each document.

Dictionary selection. The mid-left panel allows users to select or define the dictionary used for post hoc scoring. Users may either upload a custom dictionary as a comma-separated file or define a dictionary directly within the interface by entering comma-separated lemmas. This flexibility enables the use of validated lexical resources as well as ad hoc dictionaries tailored to specific research questions.

Output and visualization. Once the input data and dictionary are provided, the app generates an interactive bar chart displaying topic-level content or function scores. Each bar corresponds to a topic k, with bar length proportional to the computed score (Figure 1). Both the visualizations and the underlying numerical results can be downloaded using the interface controls.

Figure 2 – Topic TasseographeR user interface.



2.2 Measures

Topic tasseographeR computes two topic-level metrics that assist in topic interpretation and validation: the **“**topic content**”** and the **“**topic function**”** scores.

Topic content score. The *topic content score* is designed to assess post hoc the presence of specific content word dictionaries in the vector of topic-word distributions, β, which is central to the generative process of topic modelling. This metric represents a topic-level dictionary score for the construct measured by a specific dictionary.

For each topic k, β is a vector where each entry $\beta_{k,w}$ represents the probability of word w being associated with topic k [6,12]. Indeed, at the topic labelling stage [8], top words with the highest $\beta_{k,w}$ values for a topic k are commonly resorted to interpret and label that topic. For example, a topic derived from a newspaper corpus about arts-funding, whose top words are “city”, “building”, “park”, “design”, and “downtown”, might be labelled as “local government projects and funding” [5]. The *topic content score* precisely aims to assist in this step by providing a quantitative summary about how well the construct measured by a specific dictionary is reflected in a given topic.

To compute the *topic content score*, the lemmas of a predefined dictionary D not necessarily considered by the topic modelling process are first looked up in the topic-word distribution, provided by the topic modelling algorithm, using a hard-dictionary approach (i.e., providing a score of 1 if a word is matched and 0 if it is not matched). Then, the probability of word w occurring in topic k is used to weight such initial score. In LDA as well as in its alternative permutations, the topic-word distribution follows a sparse, skewed distribution (e.g., a Dirichlet prior) where only a small set of words are assigned a high probability of representing a given topic [6,12]. This ensures that high D scores are assigned only to topics whose vocabulary matches those in D. Finally, the weighted scores are aggregated by summing them at the k topic level. The computation involves the following steps:

1. Extract terms from $\beta_{k,w}$, $T = \{t1, t2, \dots, tn\}$;
2. Convert *T* into a document-feature matrix, *M*, where rows represent documents and columns represent term counts;
3. Filter *M* using the dictionary *D*, retaining only terms in the dictionary, to reduce sparsity;
4. Merge the dictionary-filtered matrix M_D with the original to produce a combined data frame;
5. Compute a new column, *B score*, as:

$$B\ score = D_{LEM} \cdot \beta \cdot 100$$

where $\boldsymbol{D_{LEM}}$ represents the counts of dictionary lemmas in each document, and β is a weighting factor for terms (i.e. the word probability provided by $\beta k,w$), expressed on a percent basis;

6. Finally, aggregate *B score* at each topic-level, by summing:

$$Content\ score(topic\ k) = \sum_{w \in D} B\ score$$

Higher scores of the topic content metric indicate that the vocabulary distinguishing a given topic *k* reflects the construct matched by the selected post hoc dictionary *D* to a greater extent than other topics.

Topic function score. The performance of topic models is heavily influenced by text pre-processing choices [8]. Indeed, especially in bag-of-words models, pre-processing steps like document cleaning, enrichment, stopword removal, word normalization, stemming and lemmatization, and focusing on specific parts of speech are typically undertaken before configuring the topic modelling algorithm to enhance the algorithm's ability to identify meaningful and intelligible topics [8]. However, these steps often exclude elements like function words (e.g., pronouns, conjunctions, negations, prepositions) and paralinguistic features (e.g., punctuation, emojis), which limits the application of the topic content metric. Such exclusions are significant because these elements can reveal relational, psychological, and structural aspects of communication [22,23]. During the topic labelling and validation phases [1,8], human coders are required to qualitatively inspect subsamples of documents with a high prevalence of a given topic, as identified by the document-topic distribution θ. This distribution, θ, represents the extent to which each topic is expressed in each document. A higher θ_k indicates a stronger presence of topic k in document d [12]. Accordingly, a topic label assigned by a researcher (e.g., labeling topic 1 as “local government projects and funding”) is considered valid if a close reading of the top documents by multiple human coders confirms that their primary themes align with the assigned label (e.g., newspaper articles discussing local government projects; [5]).

To address the limitations of the topic content metric, the topic function metric shifts its focus from the topic-word distribution to the document-topic distribution, θ. For each document d in a topic k, a weighted score, *T score*, is computed, reflecting how well the document aligns with the topic based on the presence of dictionary terms and the topic’s distribution in the document. In notation:

$$T\ score = D_{LEM} \cdot \theta\ k$$

Where D_{LEM}, as for the topic content score, is the count of terms from the predefined dictionary that appear in the document. In this case, the dictionary is used to capture specific lemmas (e.g., pronouns, negations, or paralinguistic elements) excluded during the pre-processing phase. The multiplication weights the dictionary term count by the document-topic distribution, ensuring that documents with higher relevance to the topic contribute more to the overall score. After calculating the weighted score for each document in a topic, scores are aggregated at the topic-level to provide a total score for the topic across all n documents. Mathematically:

$$Function\ score\ (topic\ k) = \sum_{doc=1}^{n} T\ score_{k,doc}$$

Higher scores of the topic function metric indicate that the documents predominantly containing a given topic *k* reflects the construct matched by the selected post hoc dictionary *D* to a greater extent than in other topics.

Overall, both the *topic content* and *topic function* scores aim to achieve the same goal: providing a quantitative post hoc summary of the presence of specific linguistic constructs within the topics generated by a topic modelling algorithm. These metrics, as we will see in the next section, can be used to enhance the labelling and validation of topic labels. The choice between the two metrics depends on the nature of the post hoc dictionary being used. If the lemmas in the dictionary are likely excluded during the topic modelling process—either because they were removed during pre-processing or because they are extremely common words—the researcher should use the topic function metric. Otherwise, the topic content metric is the more appropriate choice.

3. Illustrative cases

3.1. Case 1: Topic labelling and validation

Topic tasseographeR supports both deductive validation and abductive labelling of topic solutions. After labelling topics, researchers may formulate hypotheses regarding their linguistic or psycholinguistic properties and test these hypotheses using validated dictionaries. Conversely, dictionary scores can also support abductive reasoning by highlighting latent distinctions between topics that may otherwise be difficult to interpret.

As an illustration, using the "polyblogs2008" dataset [24], we estimated a 20-topic Structural Topic Model [12], setting topical prevalence as a function of political orientation and the posting day. Replication code for these analyses is available in the "example" directory on our Github repository. Topic 10, that we labelled "judicial processes", was hypothesized to exhibit high levels of legal contestation language. Applying the Loughran–McDonald litigiousness dictionary [25] confirmed this expectation, as Topic 10 displayed the highest content score across all topics (Figure 3). Similarly, Topic 20 was labelled "religion" because blog posts featuring this topic focused on religious events and themes. We hypothesized that this topic would be characterized by a high extent of language associated with religious matters. Using the "Relig" dictionary [26], we computed topic content scores, which again confirmed our labelling, with Topic 20 exhibiting the highest score across topics (Figure 4). Lastly, we also tested assumptions about function words that were likely removed during pre-processing. For example, topic 5 was labelled "elections", as its associated blog posts focused on electoral campaigns. Electoral rhetoric often employs "us-vs-them" framing, operationalized through the use of inclusive pronouns (e.g., "we," "our" [27]). Using the "Our" dictionary [26], entered via the "user defined" box of the topic function module of topic tasseographeR, we tested this hypothesis. The resulting topic content score distribution supported our labelling, as Topic 5 exhibited the highest score across all topics (Figure 5).

Figure 3 – Example 1 Results: topic tasseographeR's scores based on "Litigious" dictionary.

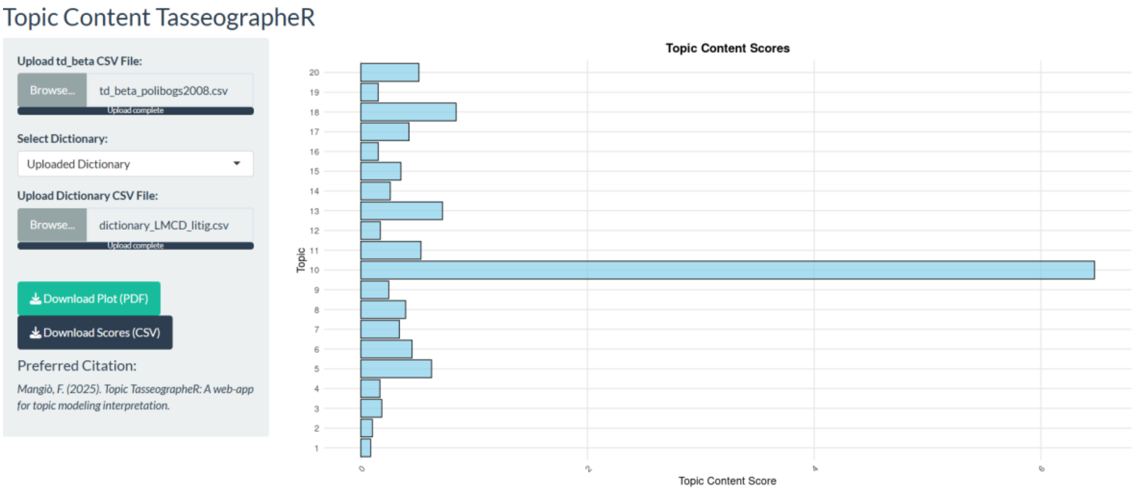


Figure 4 – Example 1 Results: topic tasseographeR's scores based on "Religion" dictionary.

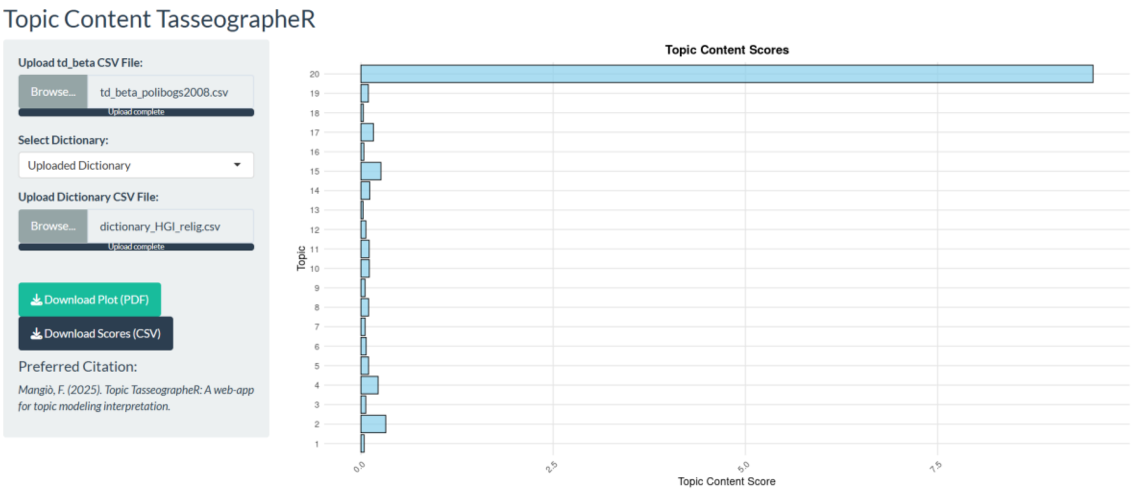
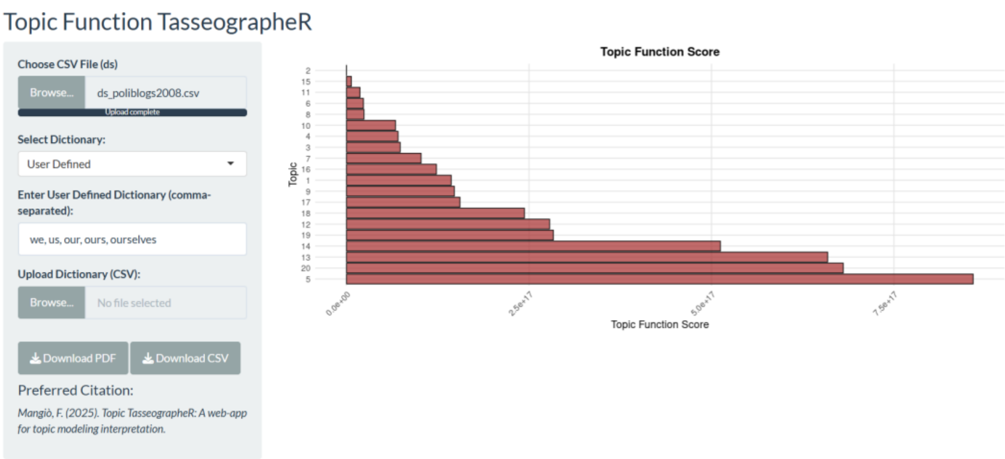


Figure 5 – Example 1 Results: topic tasseographeR's scores based on "Our" dictionary.



3.2. Case 2: Aspect-based sentiment analysis

Topic tasseographeR also enables post hoc aspect-based sentiment analysis. Whilst traditional sentiment analysis regards the computational detection of affective dimensions at the overall document-level, aspect-based sentiment analysis offers a more granular approach. It extracts specific aspects within documents and identifies their corresponding sentiment polarity from opinionated texts [28]. Several approaches merging topic and affective analysis have been so far proposed to conduct aspect-based sentiment analysis. In this vein, affective structure can be added before the computation of the models, for instance computing sentiment scores at the document level and then including such scores as topic prevalence covariates in a structural topic model [29,30]. Other approaches involve considering sentiment information during the model estimation, for example including the lemmas of a sentiment dictionary as seeds for a semi-supervised topic modelling [31,32]. While these approaches are valid and useful, they rely on a top-down sentiment measurement instrument that must be predetermined before the actual content of the topics is fully understood.

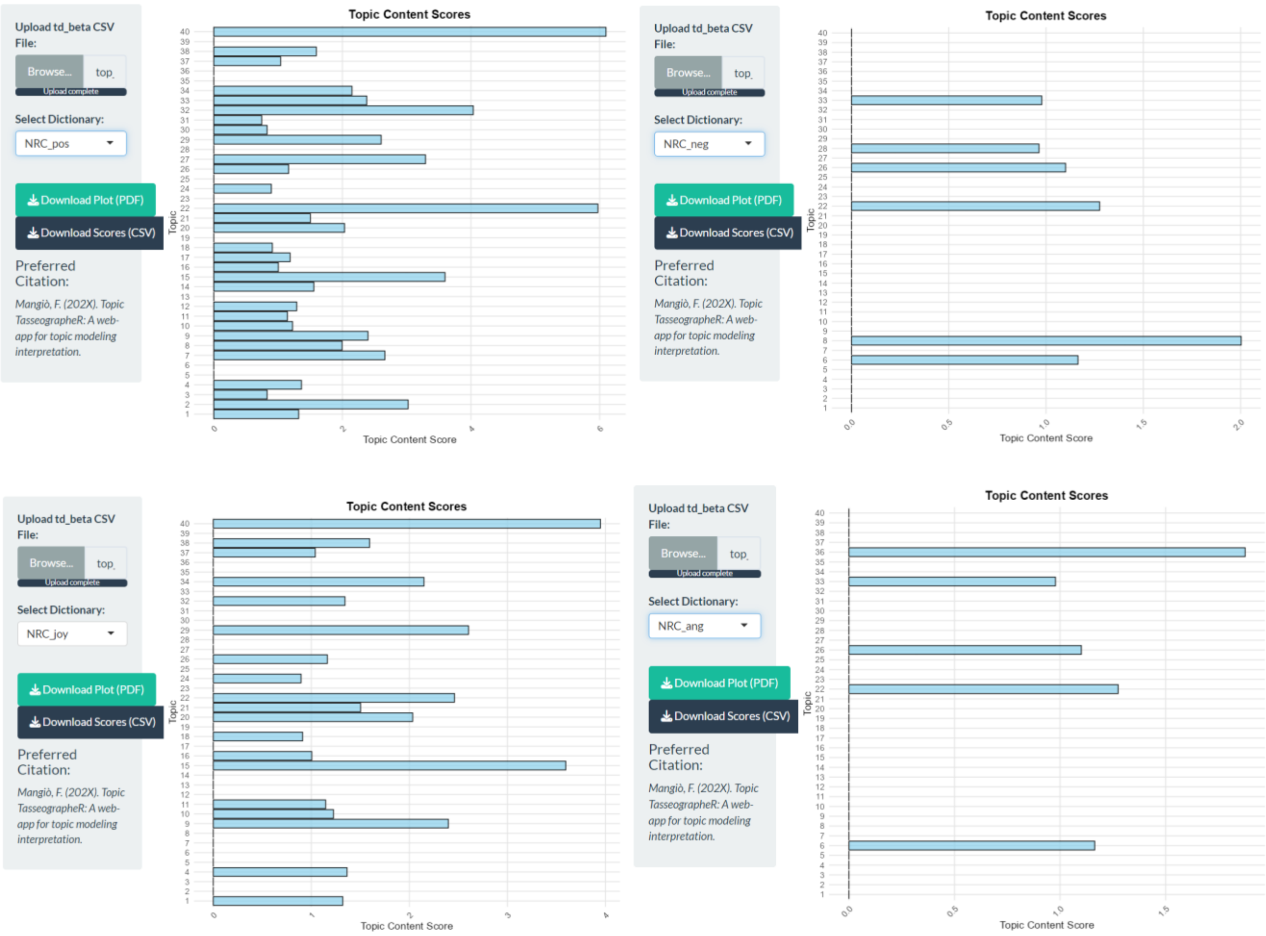
Topic tasseographeR enables an alternative, sequential approach in which the sentiment of topics is computed only after the topics have been rendered. This approach aligns with the iterative nature of text-as-data paradigms [8] and allows researchers to apply multiple sentiment dictionaries in sequence without compromising the integrity of the focal topics.

To illustrate the use of topic tasseographeR for this purpose, we conducted an aspect-based sentiment analysis on the Deceptive Opinion Spam corpus [33]. First, to perform aspect extraction [28], i.e., identifying the specific aspect elements – such as services, facilities, experiences – discussed in the hotel reviews, we estimated an LDA-based 40-topic solution using the topicmodels package [34], focusing on the truthful reviews. Replication code for these analyses is available in the “example” directory on our Github repository. Next, we proceed to assess the affective dimension of these extracted aspects. To this purpose, we loaded the resulting topic-word distribution in the topic content module of topic tasseographeR. As sentiment dictionary, we opted for the English version of the NRC Emolex. For this illustration, we focused on the positive and negative valence and, among the discrete emotions, joy and anger.

The results reveal distinct sentiment patterns associated with various topics (Figure 6). For example, a highly appreciated aspect pertains the facilities (e.g., “gym”, “bathroom”, topic 22) and the convenient location of the hotel in the Michigan area (topic 40). Indeed, this topic recorded the highest score for positive sentiment and joy, alongside the lowest score for negative sentiment. Conversely—and somewhat unsurprisingly [30]—lunch-related experiences (e.g., breakfast and restaurant services- topic 22) emerged as a polarized aspect. Topic 22 exhibited relatively high scores for both positive and negative sentiment, as well as for the discrete emotions of joy and anger (Figure 6).

This example demonstrates how topic tasseographeR can be employed to conduct aspect-based sentiment analysis, allowing researchers to effortlessly compute topic-specific sentiment scores post hoc.

Figure 6 – Example 2 Results: topic tasseographeR's scores based on NRC's positive, negative, joy, anger dictionaries.



4. Discussion

Topic tasseographeR contributes to topic-modelling research by offering a reliable, cost-effective, and reproducible approach to topic interpretation and validation grounded in dictionary-based content analysis. By leveraging validated lexical resources, the software enhances interpretability without requiring additional model estimation or opaque inference procedures.

The software improves existing research workflows by enabling systematic hypothesis testing about topic semantics and facilitating comparisons across topics, dictionaries, and corpora. It supports both exploratory and confirmatory analyses and integrates seamlessly with established topic-modeling pipelines, as evidenced by the extant research that employed this paradigm across multiple empirical contexts and social science domains [18,19,20,21].

At the same time, dictionary-based validation should not be treated as a standalone replacement for qualitative inspection or alternative validation strategies. Careful dictionary selection remains essential, particularly when applying lexical resources to corpora that differ substantially from those used during dictionary construction. Although illustrated using probabilistic topic models, the underlying methodology is transferable to a wide range of topic-modelling approaches, including CorEx [36], LSA [37], Top2Vec [38], and BERTopic [7], provided that topic–word and document–topic representations are available or can be reasonably approximated. Finally, while large language models offer promising zero-shot annotation capabilities, concerns related to interpretability, reproducibility, and cost underscore the continued relevance of transparent, dictionary-based approaches [15]. Topic tasseographeR thus provides a complementary tool that preserves researcher agency and methodological clarity in text-as-data research.

5. Conclusion

Topic tasseographeR addresses a central methodological challenge in topic modeling: the interpretation and validation of topic solutions. By combining topic-model outputs with validated lexical dictionaries, the app offers a transparent and reproducible approach to assessing topic semantics without requiring additional model estimation or opaque inference procedures. Its dual focus on topic content and topic function enables researchers to examine both lexical meaning and linguistic structure, including features often excluded during preprocessing. As a flexible, model-agnostic tool, Topic tasseographeR complements existing validation strategies and supports rigorous, theory-informed text analysis across a wide range of research contexts.

Data availability

Data, source code of the app and replication code for illustrative cases is available in the dedicated Github repository.

Funding sources

The authors have no relevant financial or non-financial interests to disclose.

References

[1] Hannigan TR, Haans RFJ, Vakili K, Tchalian H, Glaser VL, et al. Topic modeling in management research: Rendering new theory from textual data. *Acad Manag Ann* 2019;13(2):586–632. https://doi.org/10.5465/annals.2017.0099.

[2] Chen Y, Peng Z, Kim SH, Choi CW. What we can do and cannot do with topic modeling: A systematic review. *Commun Methods Meas* 2023;17(2):111–130. https://doi.org/10.1080/19312458.2023.2167965.

[3] Humphreys A, Wang RJH. Automated text analysis for consumer research. *J Consum Res* 2018;44(6):1274–1306. https://doi.org/10.1093/jcr/ucx104.

[4] Mohr JW, Bogdanov P. Introduction—Topic models: What they are and why they matter. *Poetics* 2013;41(6):545–569. https://doi.org/10.1016/j.poetic.2013.10.001.

[5] DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics* 2013;41(6):570–606. https://doi.org/10.1016/j.poetic.2013.08.004.

[6] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3(Jan):993–1022. https://doi.org/10.5555/944919.944937.

[7] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* 2022. arXiv:2203.05794.

[8] Grimmer J, Roberts ME, Stewart BM. *Text as data: A new framework for machine learning and the social sciences*. Princeton: Princeton University Press; 2022.

[9] Bernhard J, Teuffenbach M, Boomgaarden HG. Topic model validation methods and their impact on model selection and evaluation. *Comput Commun Res* 2023;5(1):1. https://doi.org/10.5117/CCR2023.1.13.BERN.

[10] Chang J, Gerrish S, Wang C, Boyd-Graber J, Blei D. Reading tea leaves: How humans interpret topic models. *Adv Neural Inf Process Syst* 2009;22.

[11] Eshima S, Imai K, Sasaki T. Keyword-assisted topic models. *Am J Polit Sci* 2024;68(2):730–750. https://doi.org/10.1111/ajps.12779.

[12] Roberts ME, Stewart BM, Tingley D. Stm: An R package for structural topic models. *J Stat Softw* 2019;91:1–40.

[13] Yang X, Zhao H, Phung D, Buntine W, Du L. LLM reading tea leaves: Automatically evaluating topic models with large language models. *Trans Assoc Comput Linguist* 2025;13:357–375. https://doi.org/10.1162/tacl_a_00744.

[14] Kozlowski D, Pradier C, Benz P. Generative AI for automatic topic labelling. *arXiv* 2024. arXiv:2408.07003.

[15] Ziems C, et al. Can large language models transform computational social science? *arXiv* 2023. https://doi.org/10.48550/arXiv.2305.03514.

[16] Baden C, Pipal C, Schoonvelde M, van der Velden MACG. Three gaps in computational text analysis methods for social sciences: A research agenda. *Commun Methods Meas* 2022;16(1):1–18. https://doi.org/10.1080/19312458.2021.2015574.

[17] Pennebaker JW. *The secret life of pronouns: What our words say about us*. New York: Bloomsbury Press; 2011.

[18] Dehler-Holland J, Okoh M, Keles D. Assessing technology legitimacy with topic models and sentiment analysis—The case of wind power in Germany. *Technol Forecast Soc Change* 2022;175:121354. https://doi.org/10.1016/j.techfore.2021.121354.

[19] Mangiò F, Pedeliento G, Andreini D, Zarantonello L. How persuasive is woke brand communication on social media? Evidence from a consumer engagement analysis on Facebook. *J Brand Manag* 2024;31(4):345–381. https://doi.org/10.1057/s41262-023-00347-4

[20] Murtas G, Mangiò F, Pedeliento G, Manoli AE. Audience perceptions of athletes' brand self-presentation on social media. *Eur Sport Manag Q* 2025;25(5):727–749. https://doi.org/10.1080/16184742.2024.2424299.

[21] Moro S, Pires G, Rita P, Cortez P. A cross-cultural case study of consumers' communications about a new technological product. *J Bus Res* 2020;121:438–447. https://doi.org/10.1016/j.jbusres.2018.08.009.

[22] Feuerriegel S, Hartmann J, Janiesch C, Zschech P. Generative AI. *Bus Inf Syst Eng* 2024;66(1):111–126. https://doi.org/10.1007/s12599-023-00834-7.

[23] Berger J, Packard G. Wisdom from words: The psychology of consumer language. *Consum Psychol Rev* 2023;6(1):3–16. https://doi.org/10.1002/arcp.1085.

[24] Warin T. Global research on coronaviruses: An R package. *J Med Internet Res* 2020;22(8):e19615. https://doi.org/10.2196/19615.

[25] Loughran T, McDonald B. The use of word lists in textual analysis. *J Behav Finance* 2015;16(1):1–11. https://doi.org/10.1080/15427560.2015.1000335.

[26] Stone PJ, Bales RF, Namenwirth JZ, Ogilvie DM. The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behav Sci* 1962;7(4):484.

[27] Holt D, Cameron D. *Cultural strategy: Using innovative ideologies to build breakthrough brands*. Oxford: Oxford University Press; 2010.

[28] Tang F, Fu L, Yao B, Xu W. Aspect-based fine-grained sentiment analysis for online reviews. *Inf Sci* 2019;488:190–204. https://doi.org/10.1016/j.ins.2019.02.064.

[29] Fresneda JE, Burnham TA, Hill CH. Structural topic modelling segmentation: A segmentation method combining latent content and customer context. *J Mark Manag* 2021;37(7–8):792–812. https://doi.org/10.1080/0267257X.2021.1880464.

[30] Hu N, Zhang T, Gao B, Bose I. What do hotel customers complain about? Text analysis using structural topic model. *Tour Manag* 2019;72:417–426. https://doi.org/10.1016/j.tourman.2019.01.002.

[31] Watanabe K. Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. *Commun Methods Meas* 2021;15(2):81–102. https://doi.org/10.1080/19312458.2020.1832976.

[32] Pipal C, Schoonvelde M, Schumacher G, Boiten M. JST and rJST: Joint estimation of sentiment and topics in textual data using a semi-supervised approach. *Commun Methods Meas* 2024:1–19. https://doi.org/10.1080/19312458.2024.2383453.

[33] Ott M, Choi Y, Cardie C, Hancock JT. Finding deceptive opinion spam by any stretch of the imagination. In: *Proc 49th Annu Meet Assoc Comput Linguist: Hum Lang Technol*; 2011.

[34] Grün B, Hornik K. *topicmodels: Topic Models*. R package version 0.2-17; 2024. https://CRAN.R-project.org/package=topicmodels.

[35] Mohammad SM, Turney PD. Crowdsourcing a word–emotion association lexicon. *Comput Intell* 2013;29(3):436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x.

[36] Gallagher RJ, Reing K, Kale D, Ver Steeg G. Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Trans Assoc Comput Linguist* 2017;5:529–542.

[37] Wild F. *lsa: An open source LSA package for R*. R package version 0.73; 2015. https://rdocumentation.org/packages/lsa/versions/0.73.3.

[38] Selivanov D, Bickel M, Wang Q. *text2vec: Modern text mining framework for R*. R package version 0.6.4; 2023. https://CRAN.R-project.org/package=text2vec.