# Identifying Coherent Topics in Multimodal Data by Leveraging Common Sense Knowledge

**Felipe González-Pizarro**
Department of Computer Science
University of British Columbia
felipegp@cs.ubc.ca

## 1 Introduction

The constant increase in the volume of textual data has led to the development of various algorithms intended to summarize and understand unstructured textual data (Peter et al., 2015). A solution to this problem is topic modeling, a statistical approach for extracting core themes or *topics* from large text corpora. Thus, when a topic modeling algorithm is applied to a large corpus of documents, such as a collection of news articles, the results might include a list of topics, such as politics, economy, or sports. Usually, these topics are described by a set of representative words or phrases ranked according to their importance for the topic (El-Assady et al., 2018).

Although powerful, topic models do not always generate understandable or useful results for humans (Bianchi et al., 2021; Harrando and Troncy, 2021). Poor quality topics are those that: (1) contain incoherent or loosely connected terms (Smith et al., 2018; Wang et al., 2019; Bianchi et al., 2021); (2) are misaligned with an expert's understanding of the domain (Smith et al., 2018), (3) or do not match the users' current information needs (Hoque and Carenini, 2015; Wang et al., 2019).

Part of this problem is because most topic modeling approaches focus on the co-occurrence of terms as the primary signal to detect the semantic relations among them (Harrando and Troncy, 2021). As a result, these algorithms do not capture semantic and lexical relations between words that are not present in the corpus (Harrando and Troncy, 2021; Song et al., 2020; Hong et al., 2020). Prior work has suggested using external knowledge to overcome this drawback (Hong et al., 2020), and common sense knowledge is one promising alternative (Harrando and Troncy, 2021). While there have been some efforts to incorporate this general human knowledge to improve the interpretability of automatically generated topics (Harrando and Troncy, 2021; Rajagopal et al., 2013), none of these algorithms supports multimodal data.

To mitigate this problem, in this paper, we propose a new topic modeling algorithm that leverages common sense knowledge to identify coherent topics in multimodal data. We train and evaluate our algorithm on a multimodal dataset of 100,000 Antisemitic/Islamophobic posts from 4chan. Our results show that injecting common sense knowledge into a topic modeling algorithm might improve the quality of topics.

We summarize prior work on common sense based topic modeling algorithms in Section 2. Section 3 details our proposed topic modeling algorithm, while Section 4 reports our evaluation mechanisms and findings. Section 5 discusses our results and provides our lessons learned, limitations and future work. Finally, Section 6 provides our conclusions.

**Ethical considerations.** We emphasize that we rely entirely on publicly available and anonymous data shared on 4chan's /pol/. We follow standard ethical guidelines (Rivers and Lewis, 2014), like reporting our results on aggregate and not attempting to deanonymize users.

**Disclaimer.** This manuscript contains Antisemitic and Islamophobic textual and graphic elements that are offensive and are likely to disturb the reader.

## 2 Common sense based topic modeling algorithms

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is one of the most popular topic modeling techniques (Meeks and Weingart, 2012; Qiang et al., 2020). It is based on the assumption that document collections have latent topics, which are typically presented to users via its *top-N* highest probability words (Lau et al., 2014). The algorithm models the documents as a bag of words to identify meaningful topics. Unfortunately, it does not represent the documents accurately when its content is short (Liu et al., 2018), lacks regular patterns (Liu et al., 2018), and the relations between some terms are not explicitly present in the training dataset. (Harrando and Troncy, 2021).

Prior work has suggested that using external knowledge, such as common sense, might help obtain a better representation of the content of documents (Shah et al., 2021). Common sense has already been used for different tasks such as question answering (Bauer et al., 2018), sentiment analysis (Ghosal et al., 2020), and dialogue (Young et al., 2018). However, only a few attempts into incorporating common sense knowledge into topic modeling algorithms exist (Rajagopal et al., 2013; Harrando and Troncy, 2021).

One of these approaches is the Common Sense Topic Model (CSTM) (Harrando and Troncy, 2021). This recently proposed topic modeling technique augments clustering with knowledge extracted from the ConceptNet (Speer et al., 2017) to find topics that are more interpretable by humans. After evaluating this approach on several datasets, the authors claim their proposal generally performs better than the traditional LDA. However, when the quality of the topics is calculated based on the coherence of the top ten terms, LDA is still performing better.

Another common sense based topic modeling approach was also proposed (Rajagopal et al., 2013). In this algorithm, every document is represented as a bag of concepts instead of the traditional bag of words. These concepts might be keywords or phrases from the corpus. The authors represent the documents as the union of the set of common sense knowledge related to each associated concept. Then, they create vector representations of these documents and cluster them using group average agglomerative clustering. An evaluation of their approach on the 20 newsgroup dataset shows that their algorithm performs better than LDA in precision, recall, and F-measure. However, the authors did not evaluate the coherence of the resulting topics, which is an important metric to identify the quality of the resulting topics (Röder et al., 2015).

All the reviewed approaches might be helpful to identify topics over textual content. However, with the proliferation of web-based social media, it is necessary to develop topic modeling algorithms that support multimodal datasets. Social media users post both textual and image data to discuss different topics (Mittos et al., 2020). These images might be valuable information that might help obtain more meaningful topics. Nevertheless, to the best of our knowledge, no attempt to incorporate common sense knowledge in multimodal topic modeling algorithms has been proposed.

## 3 Proposal

This paper proposes a new topic modeling algorithm that leverages common sense knowledge to identify coherent topics in a multimodal dataset. To do so, we implement a five-step methodology (see Figure 1). First, we retrieve Antisemitic and Islamophobic multimodal posts from 4chan (see Section 3.1). Second, we create a document representation of the content of those posts (see Section 3.2), and we extend that information by using a popular common sense knowledge base. Third, we reduce the dimensions of these documents' representations (see Section 3.3) before categorizing them into meaningful topics (see Section 3.3.1). Finally, we retrieve the most relevant keywords and most representative images for each topic.

### 3.1 Dataset

This work focuses on 4chan, particularly the Politically Incorrect board (/pol/). /pol/ is the main board for discussing world events and politics and is infamous for spreading conspiracy theories (Zannettou et al., 2017; Tuters et al., 2018) and racist/hateful content (Hine et al., 2017; Zannettou et al., 2020). We use the publicly available dataset released by (González-Pizarro and Zannettou, 2022); this dataset includes 573,513 Antisemitic/Islamophobic multimodal posts shared on 4chan in the period between July 1, 2016, and December 31, 2017.

**Why this dataset?** Social media sites such as Twitter, Facebook, and 4chan allow users to share their ideas and opinions instantly. However, there are several ill consequences, such as online harassment, trolling, cyber-bullying, fake news, and hate speech. We believe that exploring these conversations could help us understand how these communities interact on these platforms. Moreover, it is the first step before creating automated hate speech detection and content moderation systems.

### 3.2 Document representation

We represent documents as a combination of several elements: (1) their textual and image con-

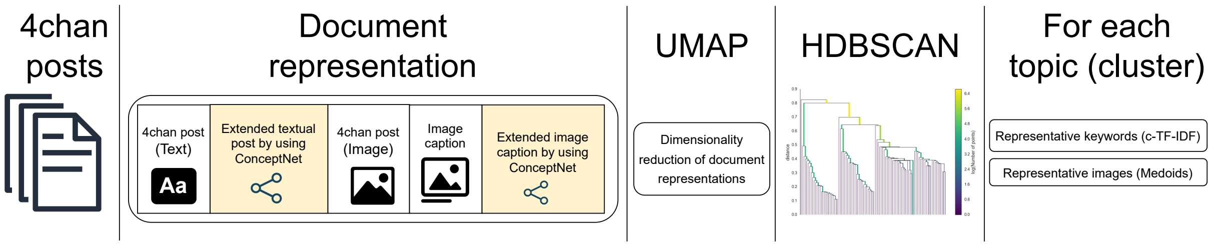

4chan posts
Document representation
4chan post (Text)
Aa
Extended textual post by using ConceptNet
4chan post (Image)
Image caption
Extended image caption by using ConceptNet
UMAP
Dimensionality reduction of document representations
HDBSCAN
For each topic (cluster)
Representative keywords (c-TF-IDF)
Representative images (Medoids)

Figure 1: Five-step methodology of our common sense based multimodal topic modeling algorithm

tent, (2) their image caption, (3) and the common sense based expansions of the textual content of the post and its image caption. We encode each element separately using the textual encoder or image encoder from the CLIP model, obtaining a high-dimensional vector for each component. The concatenation of these embeddings represents a document.

### 3.2.1 OpenAI's CLIP

OpenAI recently released a model called Contrastive Language-Image Pre-training (CLIP) (Radford et al., 2021) that leverages Contrastive Learning to generate representations across text and images. The model relies on a text encoder and an image encoder that maps text and images to a high-dimensional vector space. Subsequently, the model is trained to minimize the cosine distance between similar text/image pairs. To train CLIP, OpenAI created a vast dataset that consists of 400M pairs of text/images collected from various Web sources and covers an extensive set of visual concepts[1]. By training CLIP with this vast dataset, the model learns general visual representations and how these representations are described using natural language, which results in the model obtaining general knowledge in various topics (e.g., identifying persons, objects). In this work, we use the CLIP model to extract representations of the textual and image content of each 4chan post, its image caption, and the common sense based extensions of the textual content of the post and its image caption (see Figure 1 Document representation).

### 3.2.2 Knowledge Base

We expand the textual content of the post and its image caption based on relevant information from ConceptNet (Speer et al., 2017). This is a large-scale concept-centric knowledge base (Chakrabarty et al., 2021) that models lexical and semantic relationships (e.g., "party" *like* "flu"; "flu" *not desires* "person"). We choose ConceptNet given its vast number of concepts (approximately 1.5M nodes) and types of relations (34 in total).

### 3.2.3 Common sense based expansions

We follow several steps to get the common sense based expansions of the textual content of the posts and their image caption. First, we identify their top $j$ relevant terms by using TF-IDF (Ramos et al., 2003). Then, we retrieve the shortest path between each pair of relevant terms in our common sense knowledge base. We only select paths of length lower or equal to $k$. Finally, we convert the retrieved relations to natural language using a ConceptNet relation template[2]. Table 1 shows examples of these expansions.

### 3.2.4 Image captioning

We believe that images can provide valuable information. Thus, we use the image encoder of the CLIP model to get the vector representation of images. We also believe that we require information beyond the content of the images (Wu et al., 2021) to identify meaningful and coherent topics. We obtain this additional information by getting the image's caption and its common sense based extension.

We use ClipCap (Mokady et al., 2021) to get the caption of images. ClipCap is a CLIP-based image captioning technique that does not require additional annotations, and according to the authors, it can be applied to any data. Figure 2 shows examples of generated captions after using ClipCap in our dataset.

## 3.3 Dimensionality reduction

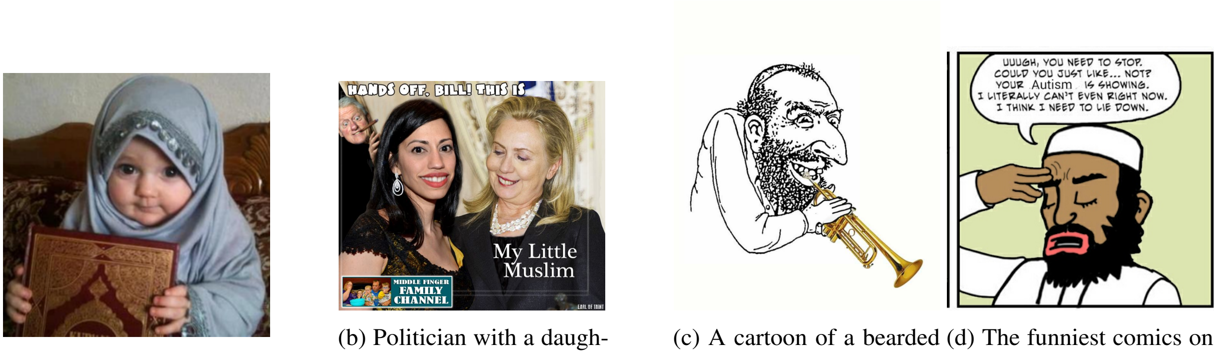
One limitation of HDBSCAN, the clustering algorithm that we choose to identify topics, is that

[1] The exact methodology for creating this dataset was not made publicly available by OpenAI.

[2] Based on https://github.com/JoshFeldman95/Extracting-CK-from-Large-LM/blob/master/templates/relation_map.json

Table 1: Sample of common based expansions. Relevant terms of the textual posts are highlighted in red

| Textual content of the post | Common sense based expansion |
|---|---|
| "I sure would. f*ck k*ke wars I refuse to fight for canada. i would fight for the usa though." | Refuse is like disagreement. Disagreement is like fight. Fight is like wars. |
| "Another white racist male trump voter" | Voter is like person. Person wants racist. Person is like male. |
| "I'm happy to accept your money and work for you" | The last thing you do when you work is get_paid. Something you might do while get_paid is happy. money makes people want work |


HANDS OFF, BILL! THIS IS
My Little Muslim
MIDDLE FINGER FAMILY CHANNEL
UUUGH, YOU NEED TO STOP. COULD YOU JUST LIKE... NOT? YOUR Autism IS SHOWING. I LITERALLY CAN'T EVEN RIGHT NOW. I THINK I NEED TO LIE DOWN.

(a) A girl with a book

(b) Politician with a daughter, and a daughter

(c) A cartoon of a bearded man playing a trumpet

(d) The funniest comics on the internet from this year

Figure 2: Examples of images captions generated by using CLIP over our dataset

its performance is reduced when there is high dimensional data. To mitigate this problem, we use Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (McInnes et al., 2018) to reduce the dimensions of the vectors of our documents' representations. This technique has been shown to preserve better the local and global features of high-dimensional data in lower projected dimensions than other traditional techniques such as PCA or t-SNE (Asyaky and Mandala, 2021; Grootendorst, 2022).

### 3.3.1 Clustering

We apply Hierarchical Density-based Spatial Clustering of Application with Noise (HDBSCAN) (McInnes and Healy, 2017) to identify the topics (clusters) of the dataset. We choose it because it can handle data with variable density, having better performance than other traditional clustering algorithms such as DBSCAN (Asyaky and Mandala, 2021).

## 3.4 Topic representation

The resulting clusters after applying HBDSCAN are the resulting topics from the dataset. In this model, a document can belongs only to one topic (cluster). We obtain its most representative terms and its most representative images for each topic.

### 3.4.1 Most relevant keywords

We use a cluster-based TF-IDF approach to identify the most relevant keywords of topics (Grootendorst, 2022). To do so, we consider all documents in a cluster as a single document by concatenating them. Then, we apply TF-IDF over this long document to identify the salient keywords. Thus, the relevance $R$ of a keyword $k$ in a topic $t$ is given by this equation:

$$R_{k,t} = kf_{k,t} * log(1 + \frac{A}{kf_k}) \quad (1)$$

Where the keyword frequency $kf_{k,t}$ represents the frequency of the keyword $k$ in a topic $t$. Then, we calculate the inverse topic frequency to measure how much information a keyword provide to a topic. We calculate it by taking the logarithm of the average number of words per topic $A$ divided by the frequency of keywords $k$ across all topics. We add one to the division within the logarithm to output only positive values. As an example, we show in Table 2 the top ten relevant keywords for the topic #188, sorted by their c-TF-IDF score.

Table 2: Top ten relevant keywords for the topic #188

| Keyword | c-TF-IDF score |
|---|---|
| Bernie | 0.0081 |
| Hillary | 0.0068 |
| Kill | 0.0037 |
| Jew | 0.0034 |
| Jewish Mouths | 0.0034 |
| Shut lying | 0.0034 |
| Lying Jewish | 0.0034 |
| Mouths | 0.0032 |
| Clinton | 0.0030 |
| Shut | 0.0028 |

### 3.4.2 Most representative Images

We are also interested in getting the representative image points for each topic. We do that by calculating the centroid/medoid of each topic. The medoid is the point in the cluster with the minimum average distance from all points in the cluster. Thus, we retrieve the $n$ image points closest to the medoid of each topic. As an example, we show in Figure 3 some of the most relevant images for the topic #188.

## 4 Evaluation and Results

This paper proposes a new topic modeling algorithm that leverages common sense knowledge to identify good-quality topics in multimodal data. One method to identify the quality of automatically generated topics is by measuring their coherence (Newman et al., 2010), which can be automatically calculated or reported by users (Efron et al., 2011; Lau et al., 2014). Topics are coherent when there are evident semantic relationships among their constituent components (e.g., keywords, documents, images) (Efron et al., 2011; Lau et al., 2014; Kaplan et al., 2010).

We evaluated our approach using a popular and well-known automatic coherence measure (Röder et al., 2015): $C_v$. This metric is based on the co-occurrence of terms. First, it retrieves the co-occurrence counts for the top words of topics using a sliding window and generates a set of vectors after calculating NPMI over these terms. Then, it measures the similarity between these vectors using cosine similarity. This metric gives a score for an entire topic model. Notice that while the *perplexity* measure has been widely used for topic models evaluation, we do not consider it because recent studies have shown that this metric is not correlated with human judgments (Xing et al., 2019).

We also evaluate our model regarding the number of topics, percentage of noise, and silhouette score. The percentage of noise indicates the amount of data that does not belong no any cluster. Therefore, users might prefer models with a lower percentage of noise. The silhouette score measures how similar an object is to its cluster compared to others. The silhouette scores range from -1 to +1. In this context, a high value indicates that documents match their topic and are poorly related to neighboring topics. While these metrics are important to identify the quality of resulting clusters, they do not provide information regarding the quality of the topics. This explains why we focus our analysis on the coherence score of the entire model.

Table 3 shows the number of topics, percentage of noise, silhouette, and coherence scores for different document representations. We evaluate our model every time we add a component to the document representation. We report our results considering a random sample of 100,000 posts from our dataset.

Our results show that adding common sense based expansions generated from the post's textual content slightly increases the coherence of the entire model. We also find that including in the document representation data related to the image content (e.g., the image itself, image caption, common sense based expansion over the image caption) decreases the coherence score.

We also observe that the number of topics increases when adding components into the document representation. Finally, the results show that adding common sense based expansions generated from the caption of images decreases the percentage of noise and increase the silhouette score.

## 5 Analysis

In this work, we explore the problem of finding coherent topics in multimodal data. We propose a new algorithm that leverages common sense knowledge to mitigate this problem. We represent the documents as a combination of several elements: their textual and image content, their image caption, and common sense based expansions generated from their textual content and image caption. We train

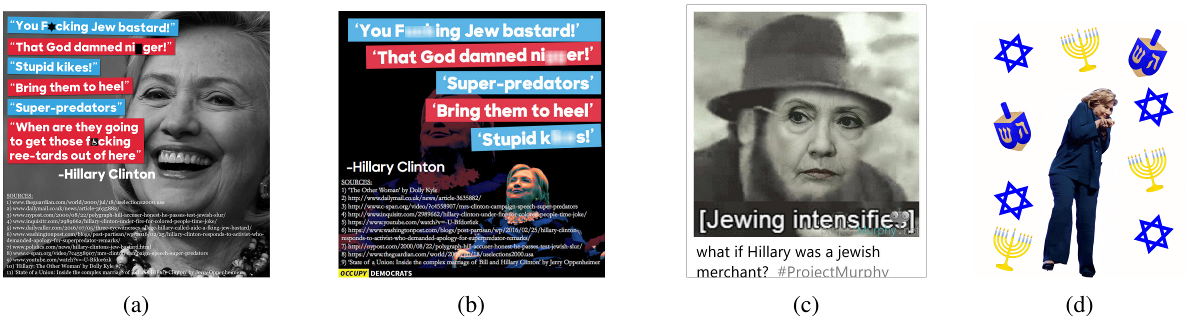

"You F[redacted]cking Jew bastard!"
"That God damned ni[redacted]ger!"
"Stupid kikes!"
"Bring them to heel"
"Super-predators"
"When are they going to get those f[redacted]cking ree-tards out of here"
-Hillary Clinton
SOURCES:
'You F[redacted]ing Jew bastard!'
'That God damned ni[redacted]er!'
'Super-predators'
'Bring them to heel'
'Stupid k[redacted]s!'
-Hillary Clinton
SOURCES:
OCCUPY DEMOCRATS
[Jewing intensifies]
what if Hillary was a jewish merchant? #ProjectMurphy
(a)
(b)
(c)
(d)

Figure 3: Top 1 (a), top 3 (b), top 8 (c) and top 10 (d) image for the topic #188

Table 3: Number of topics, percentage of noise, silhouette and coherence scores for different documents' representations. The topic model was trained over a random sample of 100,000 posts of our selected dataset. The documents' representation might include their textual content (Text), image content (Img.), image caption (Caption), and its common sense based extensions over the textual content (CS) and over its image caption (CS on caption).

| Document representation | # Topics | Noise (%) | Silhouette | Coherence |
|---|---|---|---|---|
| Text | 251 | 31.3 | 0.27 | 0.466 |
| Text + CS | 258 | 31.9 | 0.27 | 0.474 |
| Text + CS + Img. | 721 | 37.1 | 0.12 | 0.457 |
| Text + CS + Img. + Caption | 793 | 32.4 | 0.24 | 0.421 |
| Text + CS + Img. + Caption + CS on caption | 859 | 28.0 | 0.33 | 0.415 |

and evaluate our algorithm over a large-scale collection of hateful posts from the web. In this section, we discuss the potential of our approach in finding good quality topics, and we provide the lessons that we learned, limitations, and future work.

### 5.1 Common sense based expansions

Existing research work suggests that by extending the representation of documents might be possible to identify high-quality topic structures (Liu et al., 2018). Inspired by this line of work, we generate common sense based expansions and add them to the documents' representation.

Our findings show that common sense based expansions generated from the textual content of the post can slightly increase the topics' coherence. This is a relevant finding, especially because we trained the model in a non-traditional dataset. The textual post from 4chan usually contains terms and phrases (e.g., slurs) that are not present in general common sense knowledge bases (Hine et al., 2017), making it challenging to retrieve relevant information, especially when the textual content is short.

While we see the potential of using a common sense knowledge base to retrieve relevant information, we also notice that our approach is not suitable for all the cases. For example, for the phrase "Another white racist male trump voter" we obtained the following relations: "Voter is like person. Person wants racist. Person is like male". While we can expect a relation between "person" and "racist", we do not believe that they are connected by desire ("wants"). We found several similar examples while inspecting our results (e.g., "gay is the opposite of closet"). While ConceptNet contains a large number of concepts and relations, further studies should identify and remove the connections that are not correct.

We also observe that a high volume of our sentences contains the *RelatedTo* connection which our template convert it into *{a} is like {b}* where {a} and {b} are two relevant terms. Table 1 shows several examples. Indeed, one of the expansions only contains this type of relation: "Refuse is like disagreement. Disagreement is like fight. Fight is like wars". Unfortunately, this expansion is vague and does not provide more information about how the terms {"refuse", "disagreement", "fight", and "wars"} are related. We believe that by using both symmetric and asymmetric relations that provide a deeper level of description, such as *DistinctFrom*, *Causes*, *AtLocation*, the performance of our technique might improve.

### 5.2 Image captioning

Image captioning is a challenging task even today (Mokady et al., 2021) because their performance is conditioned on the datasets on which the models are trained (Zeng et al., 2022). For training ClipCap (Mokady et al., 2021), the authors used separately three different datasets: the *COCO-captions* (Common Objects in Context) (Lin et al., 2014), *nocaps* (Agrawal et al., 2019), and *Conceptual Captions* (Sharma et al., 2018). In this project, we chose the model trained on the *Conceptual Captions* (Sharma et al., 2018) dataset because it showed better qualitative results when we manually compared the captions generated for a random sample of our dataset.

We find some evidence that indicates that our chosen image captioning model can generate satisfactory captions for our selected dataset (see Figure 2 (a) ). However, it was not for all the cases. Some captions do not adequately represent the content of the images (see Figure 2 (b) ) or contain bias (see Figure 2 (d)). Prior work has already reported that captioning image models are biased based on the training data (Zeng et al., 2022). We also notice that this model needs external knowledge to generate captions more adequate to the context of the dataset. For instance, the model describes the antisemitic "Happy Merchant" meme (Zannettou et al., 2020) as "a cartoon of a bearded man playing a trumpet" (see Figure 2 (b)). This inadequate representation might impact the performance of our algorithm, mainly because these kinds of memes are prevalent in 4chan (Zannettou et al., 2018, 2020). We believe that this can explain why the coherence of the resulting topics decreases after adding to the document representation image related data.

### 5.3 Self-evaluation

In this project, we learned several lessons and dealt with several challenges. First, the initial idea was to categorize posts' textual content and image content separately. However, we discovered that while CLIP can encode textual and image content, these vector representations are significantly different. As a result, during our first attempts, the resulting topics included only textual elements or image elements, not both. That is why we decided to create a document representation that considers both modalities simultaneously. After concatenating the textual and image vector, we mitigated this problem.

We also learned that it is very computationally expensive to identify the relations between thousands of terms. Therefore, we implemented several strategies to improve the performance of the results, and all our experiments were performed in GPU clusters from the Max Planck Institute. Still, it was necessary to execute our algorithm for 10 hours and 32 minutes to identify the relevant topics from 100,000 posts.

Furthermore, while we tried to improve the common sense based expansions by (1) selecting only those with high similarity with the original document and (2) selecting a higher number of relevant terms from sentences, we had to discard those alternatives given that they largely increased the algorithm execution time.

We also identified that the selected dataset added additional complexity to our research project because common popular terms and phrases (e.g., slurs) were absent in the chosen knowledge base. However, this is not bad because we also identified room for improvement in state-of-the-art. For example, we believe that we could extend current common sense knowledge bases or create a new one to support hateful content specifically.

### 5.4 Limitations and future work

As in any study, this research has limitations that need to be considered. First, we only retrieve common sense knowledge from ConceptNet (Speer et al., 2017). We plan to extend this work by considering other common sense knowledge bases, such as Atomic (Sap et al., 2019). We expect that by considering other sources of knowledge, we can obtain more adequate common sense based expansions, especially for non-traditional datasets on which natural language processing techniques are not trained.

We also plan to compare our algorithm with other traditional topic modeling techniques such as LDA (Blei et al., 2003), or NMF (Févotte and Idier, 2011). We intend to compare these algorithms over different datasets using several automatic coherence metrics. We believe that conducting users studies will be helpful to identify the quality of the resulting topics, mainly because current automatic coherence metrics do not consider the connection between the most relevant images of topics.

We plan to optimize our current algorithm to decrease the processing time. That will allow identifying topics over larger datasets. We also intend to evaluate the performance of our approach with different parameter settings (e.g., by changing the number of relevant terms retrieved from a document, considering different lengths of the path between relevant terms in the knowledge base, measuring the similarity between the common sense based expansions and the original documents).

Our results suggest that captioning image models inject noise and decrease the performance of our algorithm. However, it is still necessary to identify the relevant elements of images. To mitigate this problem, we plan to identify the text of the images by using OCR (Memon et al., 2020) and identify the elements of images by using an object detection system (Gu et al., 2019).

## 6 Conclusions

We introduced a new topic modeling algorithm for multimodal data leveraging common sense knowledge to identify coherent topics. To the best of our knowledge, no attempt to incorporate common sense in multimodal topic modeling algorithms has been proposed. We train and evaluate our model over a large-scale collection of 4chan posts. Our results show that injecting common-sense knowledge into a topic modeling algorithm might increase topics' coherence, increase the silhouette score, and reduce the percentage of non-categorized data. Our results also show that extending the representation of documents can vastly increase the number of topics. These results hint at the potential of using external knowledge to increase the quality of topics.

## References

Harsh Agrawal, Karan Desai, Yufei Wang, Xinlei Chen, Rishabh Jain, Mark Johnson, Dhruv Batra, Devi Parikh, Stefan Lee, and Peter Anderson. 2019. Nocaps: Novel object captioning at scale. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8948–8957.

Muhammad Sidik Asyaky and Rila Mandala. 2021. Improving the performance of hdbscan on short text clustering by using word embedding and umap. In *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pages 1–6. IEEE.

Lisa Bauer, Yicheng Wang, and Mohit Bansal. 2018. Commonsense for generative multi-hop question answering tasks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4220–4230, Brussels, Belgium. Association for Computational Linguistics.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 759–766, Online. Association for Computational Linguistics.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.

Tuhin Chakrabarty, Yejin Choi, and Vered Shwartz. 2021. It's not rocket science: Interpreting figurative language in narratives. *arXiv preprint arXiv:2109.00087*.

Miles Efron, Peter Organisciak, and Katrina Fenlon. 2011. Building topic models in a federated digital library through selective document exclusion. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–10.

Mennatallah El-Assady, Fabian Sperrle, Rita Sevastjanova, Michael Sedlmair, and Daniel Keim. 2018. Ltma: Layered topic matching for the comparative exploration, evaluation, and refinement of topic modeling results. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pages 1–10. IEEE.

Cédric Févotte and Jérome Idier. 2011. Algorithms for nonnegative matrix factorization with the b-divergence. *Neural Computation*, 23(9):2421–2456.

Deepanway Ghosal, Devamanyu Hazarika, Abhinaba Roy, Navonil Majumder, Rada Mihalcea, and Soujanya Poria. 2020. KinGDOM: Knowledge-Guided DOMain Adaptation for Sentiment Analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3198–3210, Online. Association for Computational Linguistics.

Felipe González-Pizarro and Savvas Zannettou. 2022. Understanding and detecting hateful content using contrastive learning. *arXiv preprint arXiv:2201.08387*.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Yating Gu, Yantian Wang, and Yansheng Li. 2019. A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. *Applied Sciences*, 9(10):2110.

Ismail Harrando and Raphaël Troncy. 2021. Discovering interpretable topics by leveraging common sense knowledge. In *Proceedings of the 11th on Knowledge Capture Conference*, pages 265–268.

Gabriel Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. 2017. Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. In *ICWSM*.

Yang Hong, Xinhuai Tang, Tiancheng Tang, Yunlong Hu, and Jintai Tian. 2020. Enhancing topic models by incorporating explicit and implicit external knowledge. In *Asian Conference on Machine Learning*, pages 353–368. PMLR.

Enamul Hoque and Giuseppe Carenini. 2015. Convisit: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 169–180. ACM.

Ronald M Kaplan, Jill Burstein, Mary Harper, and Gerald Penn. 2010. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Jey Han Lau, David Newman, and Timothy Baldwin. 2014. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 530–539.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Chi-Yu Liu, Zheng Liu, Tao Li, and Bin Xia. 2018. Topic modeling for noisy short texts with multiple relations. In *SEKE*, pages 610–609.

Leland McInnes and John Healy. 2017. Accelerated hierarchical density based clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 33–42. IEEE.

Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Elijah Meeks and Scott B Weingart. 2012. The digital humanities contribution to topic modeling. *Journal of Digital Humanities*, 2(1):1–6.

Jamshed Memon, Maira Sami, Rizwan Ahmed Khan, and Mueen Uddin. 2020. Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr). *IEEE Access*, 8:142642–142668.

Alexandros Mittos, Savvas Zannettou, Jeremy Blackburn, and Emiliano De Cristofaro. 2020. "and we will fight for our race!" a measurement study of genetic testing conversations on reddit and 4chan. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 452–463.

Ron Mokady, Amir Hertz, and Amit H Bermano. 2021. Clipcap: Clip prefix for image captioning. *arXiv preprint arXiv:2111.09734*.

David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, pages 100–108.

Jessica Peter, Steve Szigeti, Ana Jofre, and Sara Diamond. 2015. Topicks: Visualizing complex topic models for user comprehension. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 207–208. IEEE.

Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu. 2020. Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR.

Dheeraj Rajagopal, Daniel Olsher, Erik Cambria, and Kenneth Kwok. 2013. Commonsense-based topic modeling. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '13, New York, NY, USA. Association for Computing Machinery.

Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 29–48. Citeseer.

Caitlin M Rivers and Bryan L Lewis. 2014. Ethical research standards in a world of big data. *F1000Research*, 3(38):38.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3027–3035.

Adnan Muhammad Shah, Xiangbin Yan, Samia Tariq, and Mudassar Ali. 2021. What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Information Processing & Management*, 58(3):102516.

Piyush Sharma, Nan Ding, Sebastian Goodman, and Radu Soricut. 2018. Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2556–2565.

Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system. In *23rd International Conference on Intelligent User Interfaces*, pages 293–304. ACM.

Dandan Song, Jingwen Gao, Jinhui Pang, Lejian Liao, and Lifei Qin. 2020. Knowledge base enhanced topic modeling. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 380–387. IEEE.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-first AAAI conference on artificial intelligence*.

Marc Tuters, Emilija Jokubauskaitė, and Daniel Bach. 2018. Post-truth protest: how 4chan cooked up the pizzagate bullshit. *M/c Journal*, 21(3).

Jun Wang, Changsheng Zhao, Junfu Xiang, and Kanji Uchino. 2019. Interactive topic model with enhanced interpretability. In *IUI Workshops*.

Jialin Wu, Jiasen Lu, Ashish Sabharwal, and Roozbeh Mottaghi. 2021. Multi-modal answer validation for knowledge-based vqa. *arXiv preprint arXiv:2103.12248*.

Linzi Xing, Michael J. Paul, and Giuseppe Carenini. 2019. Evaluating topic quality with posterior variability. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3471–3477, Hong Kong, China. Association for Computational Linguistics.

Tom Young, Erik Cambria, Iti Chaturvedi, Hao Zhou, Subham Biswas, and Minlie Huang. 2018. Augmenting end-to-end dialogue systems with commonsense knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the origins of memes by means of fringe web communities. In *Proceedings of the Internet Measurement Conference 2018*, pages 188–202.

Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In *IMC*.

Savvas Zannettou, Joel Finkelstein, Barry Bradlyn, and Jeremy Blackburn. 2020. A quantitative approach to understanding online antisemitism. In *ICWSM*.

Andy Zeng, Adrian Wong, Stefan Welker, Krzysztof Choromanski, Federico Tombari, Aveek Purohit, Michael Ryoo, Vikas Sindhwani, Johnny Lee, Vincent Vanhoucke, et al. 2022. Socratic models: Composing zero-shot multimodal reasoning with language. *arXiv preprint arXiv:2204.00598*.