Multimodal Classification of Violent Online Political Extremism Content with Graph Convolutional Networks

Stevan Rudinac
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
s.rudinac@uva.nl

Iva Gornishka
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
iva.gornishka@student.uva.nl

Marcel Worring
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
m.worring@uva.nl

ABSTRACT

In this paper we present a multimodal approach to categorizing user posts based on their discussion topic. To integrate heterogeneous information extracted from the posts, i.e. text, visual content and the information about user interactions with the online platform, we deploy graph convolutional networks that were recently proven effective in classification tasks on knowledge graphs. As the case study we use the analysis of violent online political extremism content, a challenging task due to a particularly high semantic level at which extremist ideas are discussed. Here we demonstrate the potential of using neural networks on graphs for classifying multimedia content and, perhaps more importantly, the effectiveness of multimedia analysis techniques in aiding the domain experts performing qualitative data analysis. Our conclusions are supported by extensive experiments on a large collection of extremist posts.

KEYWORDS

Multimedia classification; graph convolutional networks; entity linking; semantic concepts; violent online political extremism

ACM Reference format:
Stevan Rudinac, Iva Gornishka, and Marcel Worring. 2017. Multimodal Classification of Violent Online Political Extremism Content with Graph Convolutional Networks. In *Proceedings of ThematicWorkshops'17, October 23–27, 2017, Mountain View, CA, USA., ,* 8 pages.
DOI: https://doi.org/10.1145/3126686.3126776

1 INTRODUCTION

Rapid expansion of internet, content sharing and social networking platforms brought dramatic changes to our everyday communication, much of which is currently happening in the digital domain. Keeping in touch with the peers, sharing the ideas and obtaining information about practically anything has never been easier. And while it is hard to imagine a part of society that did not benefit from the "information revolution", new security challenges emerged. Internet fora and social networking platforms, for example, have provided effective means for spreading the violent online political extremism content and promoting extremist ideologies. The severity of the problem has recently motivated technology giants including Facebook, Microsoft, Twitter and YouTube to launch a joint initiative for preventing the spread of extremist content [22]. However, such projects are mostly focusing on hashing and filtering known extremist multimedia items. Aiding domain experts in rigorous large-scale empirical research and comparative analysis of the online strategies different extremist groups apply requires novel tools for multimodal analysis of data from various information sources.

Similar to conventional social multimedia, the messages exchanged in "extremosphere" typically consist of text and visual content. There is an important difference though - discussions of ideological viewpoints imply a particularly high semantic level at which the messages should be analysed. For example, the domain expert from social, political or communication sciences may be interested in knowing whether a message or a user are associated with a particular topic of interest, such as *scientific racism*, *xenophobia* or *neo-Nazism*. Additionally, aware of the fact that their expressed viewpoints may be socially unacceptable or even in collision with the law, the users of those fora may be more careful when phrasing the messages.

In this paper we present an approach to semantic categorisation of multimedia originating from violent online extremism fora. To facilitate analysis at a high semantic level (cf. Figure 1), we perform entity linking [24] in case of text and extract semantic concepts [34] from the visual content. This choice of representation produces the labels that are easily interpretable by the analyst, which makes entity linking preferable to the popular alternatives such as topic modelling [4]. Additionally, linking entities to a knowledge base like Wikipedia provides an additional context for the discussion and makes analysing messages easier. We further conjecture that the user preferences may be a good predictor of the post category and for that reason in our model we also include information about user interactions with the forum.

Graph-based approaches to multimedia information retrieval are famed for their ability to uncover hidden relations between multimedia items. However, their wider adoption is hindered by, amongst other things, the challenges associated with appropriate weighting of different modalities [8, 30]. Graph convolutional networks (GCNs) were recently introduced as a new technique with high potential for classification tasks [11, 15]. So far, their applicability was demonstrated on several different types of machine learning problems, associated with e.g. molecular fingerprinting [11], citation networks and knowledge graphs [15].

Here we investigate the potential of GCNs [15] for analysing content-rich multimedia collections. As the input into GCN we use

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ThematicWorkshops'17, October 23–27, 2017, Mountain View, CA, USA.
© 2017 ACM. ISBN 978-1-4503-5416-5/17/10…$15.00
DOI: https://doi.org/10.1145/3126686.3126776

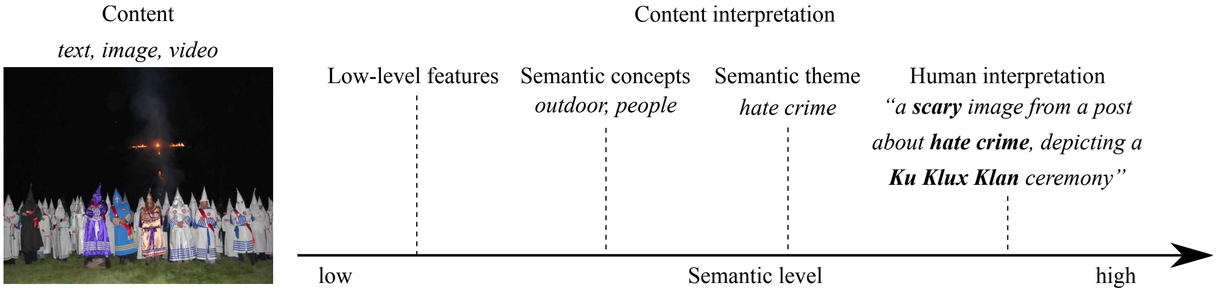


Figure 1: Semantic levels at which the content may be analysed.

a graph, constructed by treating blog posts as the nodes and their multimodal relations as the edges, as well as the features of the nodes. Next to the overall ability of our pipeline to assign blog posts to correct categories, we are interested in the contribution of each modality to the classification performance. Finally, we showcase the applicability of semantic multimedia analysis for aiding the domain experts involved in empirical research of violent online political extremism. As the test bed for our study we use a large collection of multimedia posts gathered from Stormfront, a white nationalist, white supremacist and neo-Nazi Internet forum [3]. Our choice of collection and the task is motivated by the research questions raised by the domain experts from social sciences fields as well as the forum's relatively clear structure that allows evaluation of our approach.

The following are the main contributions of this paper:

- We demonstrate potential of graph convolutional neural networks for classification tasks in challenging, content rich multimedia collections.
- Our experiments provide insights into the usefulness of individual modalities and semantic features input into the network for discriminating between the posts at the topical level.
- Our case study shows viability of multimedia analysis techniques for aiding the domain experts involved in the qualitative analysis of violent online political extremism.

The reminder of this paper is organised as follows. In Section 2 we provide an overview of related work. Then in Section 3 we introduce our approach and in sections 4 and 5 we present the experimental results. Section 6 concludes the paper.

2 RELATED WORK

Mining high-level semantics from multimedia

While the primarily focus of multimedia and computer vision communities traditionally rarely went beyond the semantic level of concepts [16], actions [33] and events [13], in the recent years facilitating multimedia information retrieval at a higher semantic level started gaining popularity. For example, Rudinac et al. utilize spoken content and semantic concept detection for video search based on semantic themes, such as politics, science, archaeology and cultural identity [31]. The authors expand topical queries using multiple query expansion techniques and use query performance prediction to identify the best results list. In [36] the authors propose an approach for organising video search results into hierarchies for easier topic-based exploration. For that they match extracted named entities (i.e. personages and locations) to the Wikipedia hierarchy, which is further adapted to the properties of retrieved videos. Another group of approaches successfully adapt proven semantic analysis techniques from the information retrieval domain, such as LDA [4], Word2vec [20] and DocNADE [17] for use with multimodal content. In the early examples [2, 28], the conventional LDA was adapted to jointly learn distribution of textual and visual clues for image/video annotation. Although the labels were defined at the level of semantic concepts, we mention them here because we believe that such approaches could be extended to multimedia categorisation at the level of semantic themes. More recently Kottur et al. extended Word2vec to learn visually-grounded word embeddings, while Zheng et al. introduced a multi-modal topic-modelling approach [39] based on DocNADE. Finally, in another interesting work Qian et al. proposed a multimodal approach for topic opinion mining from social multimedia streams [29]. Again, although not directly relevant to our approach, we mention it here due to a particularly high semantic level at which the relevance criteria are defined. To the best of our knowledge, so far there have been no works on categorizing multimedia at the level of extremist or fine-grained political ideologies in general.

Graph-based approaches to multimedia analysis

The effectiveness of graph-based approaches for uncovering hidden relations between multimedia items was time and again proven in multimedia and computer vision communities. An early example is the approach proposed by Pan et al., which jointly models regions automatically segmented from an image and the image-level annotations [27]. To determine affinities between nodes in the graph, random walk with restarts is applied, which gained a wider popularity as an integral part of PageRank algorithm [25]. The approach performed well in image auto-captioning (i.e. tag propagation) and finding correlations between different modalities. Clements et al. utilize a similar graph topology for content recommendation and personalised search in social media collections [8]. The authors stress importance of modality weighting and investigate optimal edge weights for different information access tasks. Similarly, Rudinac et al. deploy such multi-layer graph, integrating text, visual and user modalities for visual summarization of geographic areas [30]. The authors further make an attempt to automatically determine modality-dependent edge weights. Although the above-mentioned approaches managed to push forward state of the art in different multimedia information retrieval tasks, scalability and modality-dependent edge weighting in content-rich multimedia collections remain a serious challenge.

More recently, Schinas et al. proposed a graph-based approach to event detection and summarization in large social media collections [32]. A sliding time window is used to select a small subset of candidate images and construct a multi-modal graph representing their relations in different modalities. Finally, clustering is applied for event detection and tracking. Following a different approach, Yoshida et al. apply circular propagation [37] to combine graphs constructed separately from text and visual content and use it for video search reranking [38]. Another large family of approaches, successfully deploys hypergraphs for modality fusion in various applications, such as music recommendation [7], image retrieval [40] and visual summarization [35].

Despite their revolutionary impact on the approaches for extracting semantics from text [20] and visual content [16], deep convolutional networks have yet to find their way into content-rich graphs typical for multimedia applications. However, recent progress in the field gives promise that this may change. Namely, Duvenaud et al. introduce a neural network with a convolution-alike propagation rule that operates directly on graphs and successfully deploy it for molecular fingerprinting [11]. Similarly, spectral graph convolutional neural networks, originally proposed in [6] and extended in [9] were proven effective in classification of handwritten digits and news texts. Finally, Graph Convolutional Networks (GCNs) proposed by Kipf et al. demonstrated a good performance on large knowledge and citation graphs [15] and we believe that the concept could be applied to multimedia graphs too.

3 APPROACH

Given a multimodal post, our goal is to assign it to one of the predefined categories. We implement the classifier using a graph convolutional network, which takes as input an adjacency matrix of the graph, encoding the relations between the posts as well as the features of the individual posts (graph nodes). Our framework assumes that the labels (categories) are known for only certain number of nodes. The output of the classifier are the predicted categories of the posts. The pipeline of our classification approach is illustrated in Figure 2 and consists of the following steps: (1) Content representation, (2) k-nearest neighbour computation, (3) graph construction and (4) classification with graph convolutional networks. Below we describe each step in turn.

3.1 Content representation

We conjecture that the useful clues about category of the post may be extracted from its text, visual content and the information about user interactions with the forum. For that reason, next to extracting standard text features (i.e., term frequencies), for each post we apply the following analyses.

3.1.1 Semantic Concept Detection. For each image in the collection we extract 346 TRECVID semantic concepts following a pipeline described in [34]. TRECVID is a yearly benchmark organised since 2003 by the National Institute of Standards and Technology (NIST) and focusing on different aspects of content-based retrieval and exploration of digital video. The list of semantic concepts used for evaluating performance of participating systems includes those related to intelligence and security applications. Although we initially experimented with 15.000 ImageNet concepts [10] as well, qualitative analyses performed by the domain experts suggest the benefits of using TRECVID concepts for analysing violent online political extremism content, which is why we here report those results only. We start from a large collection of videos annotated at the level of semantic concepts (e.g., objects, personages, settings and events) depicted in them and train a deep convolutional neural network [16] to detect presence of those concepts in the unseen images.

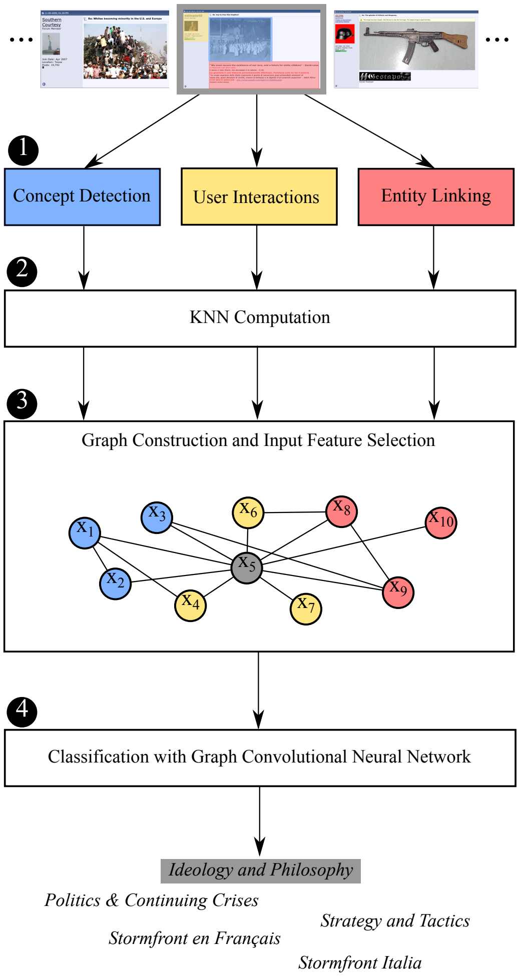


Figure 2: Pipeline of our classification approach.

3.1.2 Semantic Linking. Frame analysis is a commonly applied technique in the analysis of violent online political extremism. Domain experts usually start by manually identifying entities mentioned in the text, including the topics, people, organisations and locations. To replicate the process and facilitate semantic analysis on a much larger scale, we resort to semantic linking, where the idea is to link the text to an external knowledge base such as Wikipedia [21]. Compared to the alternatives, such as topic modeling [4], the results of semantic linking are normally easier to interpret and provide additional context for the analysed text. We link text of the Stormfront posts to English Wikipedia articles using Semanticizer [23] that was proven effective in entity linking on microblogs and conversational speech [19, 24]. The process resulted in 65.240 entities (i.e. Wikipedia articles), of which 24.929 appear only once.

3.1.3 User Relations. Stormfront platform does not support explicit friendship or follower-followee relations commonly seen in conventional social networks. Therefore, similar to [30], we model user relationships based on their topical preferences. To that end, as a measure of similarity between two users we use a relative number of messages they posted to the same forum thread (sub-category). Feature vector representing a particular user then consists of the similarities with the other users.

3.2 K-Nearest Neighbour Computation

Once the features are extracted, we proceed by computing k-nearest neighbours for each post across different modalities. Given a large dimensionality of feature vectors (e.g., 65.240 in case of entities), before k-NN computation we apply principal component analysis (PCA) to reduce them. With the exception of TRECVID concepts, our features are very sparse, which makes reduction to a small number of components reasonable.

3.3 Graph Construction and Input Features

Adding an edge between each pair of nodes for every modality would be one way to construct the graph. To that end, related work in multimedia community explored both "multi-layer" [26, 30] and multi-edge [18] graphs. However, the size of our dataset would yield prohibitively large, dense adjacency matrices. For that reason, we introduce an edge between a node and it's k-nearest neighbours in each modality. In our first experiment with the graph convolutional networks we choose not to apply separate weights for the modalities. Further, we perform an early fusion by concatenating feature vectors, reduced as described in Section 3.2 and feed them as an input into the GCN. In Section 5.2 we evaluate contribution of each modality to the classification performance.

3.4 Classification With GCN

To classify the posts, we make use of a two-layer GCN architecture described in detail in [15]. For improved readability, here we explain the most important properties. The forward model is given as follows:

$$Z = f(X, A) = \mathrm{softmax}\left(\hat{A}\,\mathrm{ReLU}\left(\hat{A}XW^{(0)}\right)W^{(1)}\right) \tag{1}$$

In Equation 1 $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, $\tilde{A} = A + I_N$ is the adjacency matrix of the graph with added self connections and I_N is the identity matrix. Further, $\tilde{D}_{jj} = \sum_j \tilde{A}_{ij}$ and $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ is a rectified linear unit. Finally, $W^{(0)}$ and $W^{(1)}$ are the layer-specific neural network weights trained using gradient descent.

The GCN is implemented in TensorFlow [1] and has a linear complexity with regard to the number of edges. As the default we use cross-entropy error over all labelled examples. However, our experiments show that a significant dis-balance between the number of posts in each category, reflects in a strong bias towards the majority classes. For that reason in Section 5 we experiment with custom loss functions applying class-specific weighting.

4 EXPERIMENTAL SETUP

4.1 Data Collection

As the testbed for the study we use Stormfront, a white nationalist, white supremacist and neo-Nazi Internet forum. The forum contains 40 high-level categories, indicating topics of discussion, ranging from "Politics & Continuing Crises", "Strategy and Tactics" and "Ideology and Philosophy" to the topics relevant to national chapters, e.g. "Stormfront en Français" and "Stormfront en Español y Portugués". We systematically crawled user posts from different sub-fora, typically consisting of text messages and images. The initial collection consisted of more than 2 million user posts and 87.000 images associated with them. The inspection of the initial collection reveiled that more than 800.000 posts belonged to a generic "Opposing Views" category (sub-forum), while the "Guidelines for Posting" category contained only a single message. After removing these two categories, for the classification experiments we arrived at the collection of more than 1.2 million posts.

Further, Stormfront users can upload a small avatar image and typically larger profile picture. Although they do not have to be the same, avatar is often a smaller, possibly cropped version of profile picture. We collected all avatars and profile pictures hosted on Stormfront, together with a basic information about the users. The resulting dataset contains more than 26.000 avatars and 21.000 profile pictures, with approximately 10.000 users having both. Despite their relatively small size, which makes identifying depicted semantic concepts a challenging task even for the humans, after a qualitative analysis of the collection and based on related work in the field, we chose to use only avatars for the study, as they may better capture personality and preferences of the users.

4.2 GCN Settings

As discussed in sections 3.2 and 3.3, we reduce computational complexity by applying several dimensionality reduction techniques. First, we apply PCA to reduce entity, term frequency and user features to 64 components. Given the lower dimensionality of semantic concept vectors, we conjecture that a lower number of principal components would suffice, so we reduce them to 32. When creating the graph, we introduce edges between a node and its 5 nearest neighbours computed independently using cosine similarity for each feature type. For the testing we use roughly 10% of data, 1% for validation and the rest for training. To mitigate the effect of a high imbalance between the number of items per category, we modify the cross-entropy loss such that the contribution of each class is reflected equally. We set the learning rate to 0.1 and train for 200 epochs with early stopping using Adam optimizer [14].

Table 1: Performance of different variants of our classification approach and the baseline.

Approach	Precision	Accuracy	Recall	F1
MLP	0.2067	0.2136	0.2374	0.1593
GCN-T	0.2351	0.2615	0.2923	0.1982
GCN-TE	0.2419	0.2629	0.2894	0.1991
GCN-TC	0.2473	0.2635	0.2905	0.2050
GCN-TU	0.2717	**0.3344**	**0.3560**	**0.2421**
GCN-TECU	**0.2766**	0.3226	0.3440	0.2362

5 EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our approach we conduct a set of experiments organised around the following questions:

(1) Can GCNs be deployed for semantically challenging classification tasks in content-rich multimedia collections?
(2) What is the added value of text, visual and user modalities for discriminating between posts based on their category?
(3) Can multimedia information retrieval techniques aid qualitative research of political extremism content?

5.1 General Classification Performance

In this section we evaluate overall performance of our multi-class classifier. Table 1 shows the results for the algorithm variant named GCN-TECU, in which all features are used for both creation of the adjacency matrix and as an input into the network. Similar to [15], as the baseline we use multi-layer perceptron (MLP). To get a better insight into the properties of compared approaches, next to accuracy, which is the metric used in the original paper introducing the GCNs, we report precision, recall and F1 measure as well. Our proposed approach clearly outperforms baseline by a significant margin with regard to all four metrics.

To dissect per-class performance, in Figure 3 we show confusion matrix yielded by our GCN-TECU approach. As discussed in Section 4.2, we intentionally choose a setting which minimizes bias towards large, general and uninformative categories such as "General Questions and Comments". We observe that our approach is particularly good at classifying posts exchanged within national chapters of Stormfront, such as France, Spain and Portugal and Italy. This is somewhat expected, as the national chapters are characterised by a certain number of topics, more frequent use of national (i.e. non-English) language, and a closed group of users discussing the matters of regional relevance. Confusion matrix further reveals that the miss-classifications are in many cases meaningful. For example, the posts from the sub-forum "Stormfront Srbija" are often miss-classified as belonging to "Stormfront Hungary" - a neighbouring country's chapter, "Stormfront Russia" - possibly due to Slavic, Orthodox and historic ties, "Stormfront Croatia" - due to mutually intelligible languages and complex relations in the recent history, and "Stormfront Europe". The categories "Politics and Continuing Crises" and "Ideology and Philosophy", facilitating similar types of ideological discussions seem to be frequently confused. Finally, the category "For Stormfront Ladies Only" focusing on the issues of relevance for female members and unique with regard to demographics and the topics being discussed, belongs to the categories recognised with an above average accuracy. The category is further frequently confused with "Dating Advice", characterised by related discussion topics and user demographics. On the example of this category in Section 5.3 we showcase the use of multimedia analytics in performing the qualitative analysis.

Table 2: List of TRECVID concepts most important for discriminating between "For Stormfront Ladies Only" sub-forum and the general Stormfront population.

#	concept name	#	concept name
1	Cats	14	Animation_Cartoon
2	Female-Human-Face-Close	15	Adult_Female_Human
3	Flowers	16	Girl
4	Dresses	17	Female_News_Subject
5	Female_Human_Face	18	Infants
6	Eukaryotic_Organism	19	Human_Young_Adult
7	Carnivore	20	Teenagers
8	Child	21	Dresses_Of_Women
9	Indian_Person	22	Two_People
10	Clearing	23	Rocky_Ground
11	Amateur_Video	24	First_Lady
12	Baby	25	Still_Image
13	Domesticated_Animal	26	Female_Person

Table 3: List of Wikipedia entities most important for discriminating between "For Stormfront Ladies Only" sub-forum and the general Stormfront population.

#	Wikipedia entity	#	Wikipedia entity
1	Race and ethnicity in the United States Census	14	Şile
2	Gendèr	15	Dôn
3	Jews	16	Cougar
4	Hello Ladies	17	Swastika
5	Múm	18	Mein Kampf
6	Race traitor	19	White nationalism
7	Lycia	20	Yōga (art)
8	White guilt	21	Hate crime
9	Étaín	22	Mestizo
10	David Duke	23	Adolf Hitler
11	Pre-eclampsia	24	Iran
12	White pride	25	Don Black (lyricist)
13	Tanning bed	26	Valley girl

5.2 Modality Analysis

We further investigate contribution of individual modalities to the performance of our classification approach. In the experiments we vary the features used for creating the adjacency matrix, but use them all as features of the nodes input into the network. Since most posts contain some text, we use term frequencies as a basis for

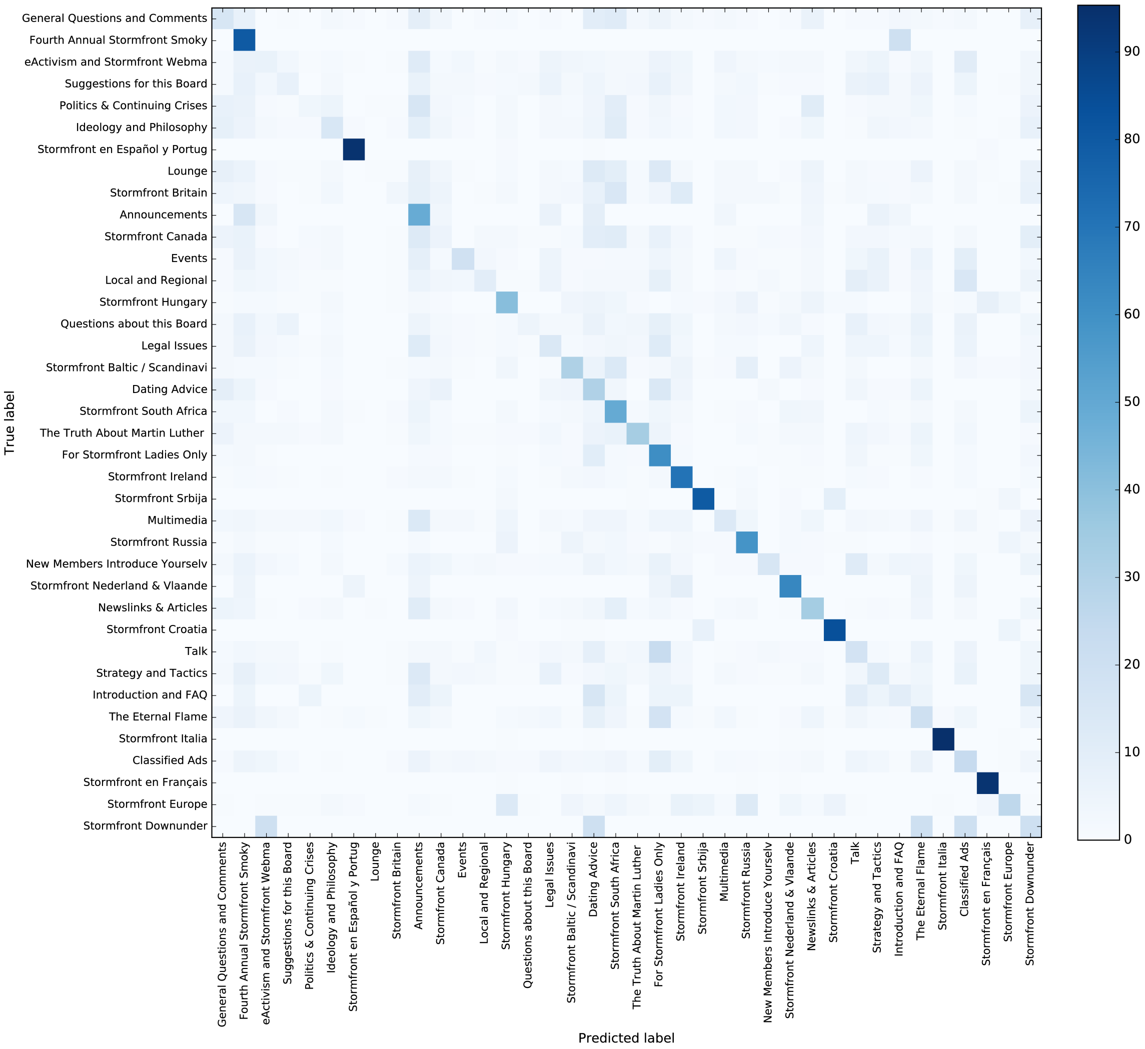


Figure 3: Confusion matrix of our GCN-TECU approach.

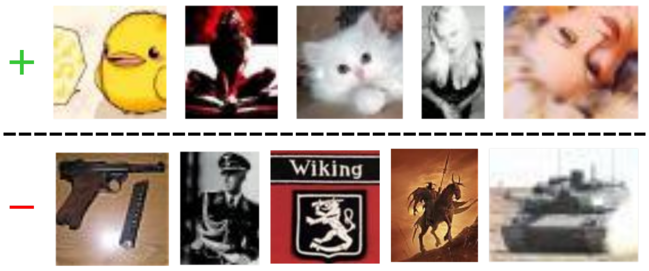
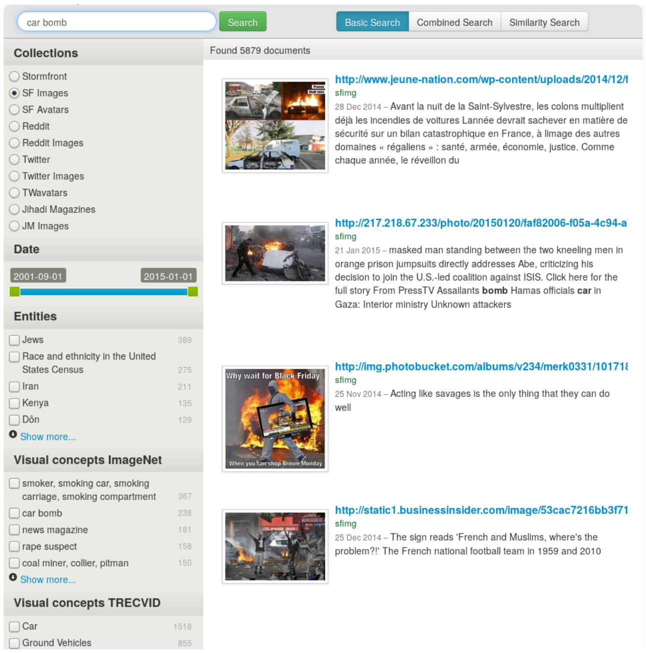


Figure 4: Example avatars selected by the users contributing to the "For Stormfront Ladies Only" sub-forum (top row) and the general Stormfront population (bottom row).

building the adjacency matrix, a setting which we denote as GCN-T. Then, we add entities (GCN-TE), semantic concepts (GCN-TC) and user features (GCN-TU). The results shown in Table 1 suggest that using a graph leads to a significant performance improvement as compared to MLP. Adding the semantic features further improves performance with regard to most considered metrics. The GCN-TC consistently outperforms GCN-TE, probably due to visual content encoding complementary information to text. Additionally, although the domain experts find them invaluable when exploring multimedia collections (cf. Figure 5), the entities are relatively sparse due to the fact that majority of posts are relatively short and characterized by the use of informal language. Utilizing user features (GCN-TU) in the creation of adjacency matrix yields a massive performance boost, which is consistent with related work on conventional graph-based approaches [30]. This is not surprising as the users normally have a limited set of interests, which makes information extracted from their profile a powerful predictor. We observe that the use of all features in creation of the adjacency matrix (i.e. GCN-TECU) improves performance of GCN-TU with regard to precision but even deteriorates it according to the other



Found 5879 documents

http://www.jeune-nation.com/wp-content/uploads/2014/12/f
sfimg
28 Dec 2014 – Avant la nuit de la Saint-Sylvestre, les colons multiplient déjà les incendies de voitures Lannée devrait sachever en matière de sécurité sur un bilan catastrophique en France, à limage des autres domaines « régaliens » : santé, armée, économie, justice. Comme chaque année, le réveillon du

http://217.218.67.233/photo/20150120/faf82006-f05a-4c94-a
sfimg
21 Jan 2015 – masked man standing between the two kneeling men in orange prison jumpsuits directly addresses Abe, criticizing his decision to join the U.S.-led coalition against ISIS. Click here for the full story From PressTV Assailants **bomb** Hamas officials **car** in Gaza: Interior ministry Unknown attackers

http://img.photobucket.com/albums/v234/merk0331/10171
sfimg
25 Nov 2014 – Acting like savages is the only thing that they can do well

http://static1.businessinsider.com/image/53cac7216bb3f71
sfimg
25 Dec 2014 – The sign reads 'French and Muslims, where's the problem?!' The French national football team in 1959 and 2010

Figure 5: Screenshot of the aggregated search system used to explore the collections and verify the results.

metrics. This is likely due to semantically related messages being posted to different categories. Therefore, we do not consider it necessarily as a drawback, but rather a useful property that could help analyst estimate seriousness of a potential threat.

5.3 Concept and Entity Importance

In this experiment we show-case viability of our chosen features, i.e. semantic concepts and the entities, for aiding the domain expert in comparative analysis of extremism content. We aim at identifying the properties of avatars specific of a particular user category. We first investigate which semantic concepts are more commonly appearing in "For Stormfront Ladies Only" forum as compared to the other fora on Stormfront. The particular question came from the domain experts investigating the role and portrayal of women in right-wing extremist networks. For that purpose, we split the collection of avatars into two classes, the first containing avatars of all users that posted at least once to the mentioned category/sub-forum and the second containing the avatars of all other users. Several examples of both categories are shown in Figure 4. We further train an extra-trees classifier [12] to evaluate the importance of each semantic concept for discriminating between the two avatar categories.

The list of top-ranked TRECVID semantic concepts, sorted by their importance in discriminating between the avatars of users posting to the "For Stormfront Ladies Only" sub-forum and the avatars of all other users is shown in Table 2. The results clearly suggest that the semantic concepts traditionally associated with the femininity are more prominent within the "Ladies" sub-forum than in general Stormfront population. The examples include semantic concepts related to pets, female face close-ups, feminine clothing, flowers, babies and infants. The conclusions were confirmed by the colleagues/project partners from the social sciences fields. To facilitate easier exploration of the collections and verify the results, we index the data in an aggregated search system Comerda [5], shown in Figure 5.

We repeat the experiment using the entities as features. The results shown in Table 3 suggest that the topics discriminating "Ladies" sub-forum from the general Stormfront population include a mix of racial subjects, references to music, television and vacation spots, as well as the references to medical and consmetic treatments and slang terms for women. Apparent is also a relative absence of political topics underlying a significant portion of discussions amongst general forum population.

6 CONCLUSION

In this paper we investigated the potential of graph convolutional networks for classification tasks in content-rich multimedia collections. As a test bed for our study we use the analysis of violent online political extremism content. The experiments conducted on a realistic collection and centred around research questions raised by the domain experts demonstrate the effectiveness of our classification approach in discriminating between user posts based on the ideas they convey. We further analysed usefulness of text, visual and user modalities for multimedia classification based on the relevance criteria specified at a particularly high semantic level. The results confirm the merit of multimodal approach and suggest that in case of limited computational resources, user modality should be preferred together with simple term frequency features. Finally, on an example case study we demonstrated that multimedia analysis techniques may be a valuable asset to domain experts performing qualitative analysis of violent online political extremism.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 312827 (NoE VOX-Pol).

The authors would like to thank Stefan Verbruggen and Gijs Koot from TNO for their help with collecting the data as well as Prof.dr. Maura Conway from the Dublin City University for fruitful discussions about the qualitative analysis of violent online political extremism content.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, and others. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).

[2] Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei, and Michael I. Jordan. 2003. Matching Words and Pictures. *J. Mach. Learn. Res.* 3 (March 2003), 1107–1135.

[3] Don Black. 1996-2017. Stormfront - a white nationalist, white supremacist and neo-Nazi Internet forum. https://www.stormfront.org/. (1996-2017). Online. Accessed on March, 2017.

[4] David M. Blei. 2012. Probabilistic Topic Models. *Commun. ACM* 55, 4 (April 2012), 77–84.

[5] Marc Bron, Jasmijn van Gorp, Frank Nack, Lotte Belice Baltussen, and Maarten de Rijke. 2013. Aggregated Search Interface Preferences in Multi-session Search Tasks. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, New York, NY, USA, 123–132.

[6] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2013. Spectral Networks and Locally Connected Networks on Graphs. *CoRR* abs/1312.6203 (2013). http://arxiv.org/abs/1312.6203

[7] Jiajun Bu, Shulong Tan, Chun Chen, Can Wang, Hao Wu, Lijun Zhang, and Xiaofei He. 2010. Music Recommendation by Unified Hypergraph: Combining Social Media Information and Music Content. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 391–400.

[8] Maarten Clements, Arjen P De Vries, and Marcel JT Reinders. 2010. The task-dependent effect of tags and ratings on social media access. *ACM Transactions on Information Systems (TOIS)* 28, 4 (2010), 21.

[9] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.). Curran Associates, Inc., 3844–3852.

[10] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.

[11] David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P. Adams. 2015. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*. MIT Press, Cambridge, MA, USA, 2224–2232.

[12] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine Learning* 63, 1 (2006), 3–42.

[13] Amirhossein Habibian, Thomas Mensink, and Cees G.M. Snoek. 2014. VideoStory: A New Multimedia Embedding for Few-Example Recognition and Translation of Events. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 17–26.

[14] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). http://arxiv.org/abs/1412.6980 Published as a conference paper at ICLR 2015.

[15] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016). Published as a conference paper at ICLR 2017.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 1097–1105.

[17] Hugo Larochelle and Stanislas Lauly. 2012. A Neural Autoregressive Topic Model. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 2708–2716.

[18] Dong Liu, Shuicheng Yan, Yong Rui, and Hong-Jiang Zhang. 2010. Unified Tag Analysis with Multi-edge Graph. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 25–34.

[19] Edgar Meij, Wouter Weerkamp, and Maarten de Rijke. 2012. Adding Semantics to Microblog Posts. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 563–572.

[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 3111–3119.

[21] David Milne and Ian H. Witten. 2008. Learning to Link with Wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*. ACM, New York, NY, USA, 509–518.

[22] Facebook Newsroom. 2016. Partnering to Help Curb Spread of Online Terrorist Content. http://newsroom.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/. (2016). Online. Accessed on March, 2017.

[23] Daan Odijk. 2012. UvA Semanticizer Web API. https://github.com/semanticize/semanticizer. (2012).

[24] Daan Odijk, Edgar Meij, and Maarten de Rijke. 2013. Feeding the Second Screen: Semantic Linking Based on Subtitles. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval (OAIR '13)*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, France, France, 9–16.

[25] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[26] Jia-Yu Pan, Hyung-Jeong Yang, Christos Faloutsos, and Pinar Duygulu. 2004. Automatic multimedia cross-modal correlation discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 653–658.

[27] Jia-Yu Pan, Hyung-Jeong Yang, C. Faloutsos, and P. Duygulu. 2004. GCap: Graph-based Automatic Image Captioning. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*. 146–146.

[28] D. Putthividhy, H. T. Attias, and S. S. Nagarajan. 2010. Topic regression multi-modal Latent Dirichlet Allocation for image annotation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3408–3415.

[29] Shengsheng Qian, Tianzhu Zhang, and Changsheng Xu. 2016. Multi-modal Multi-view Topic-opinion Mining for Social Event Analysis. In *Proceedings of the 2016 ACM on Multimedia Conference (MM '16)*. ACM, New York, NY, USA, 2–11.

[30] Stevan Rudinac, Alan Hanjalic, and Martha Larson. 2013. Generating visual summaries of geographic areas using community-contributed images. *IEEE Transactions on Multimedia* 15, 4 (2013), 921–932.

[31] Stevan Rudinac, Martha Larson, and Alan Hanjalic. 2012. Leveraging visual concepts and query performance prediction for semantic-theme-based video retrieval. *International Journal of Multimedia Information Retrieval* 1, 4 (2012), 263–280.

[32] Manos Schinas, Symeon Papadopoulos, Georgios Petkos, Yiannis Kompatsiaris, and Pericles A. Mitkas. 2015. Multimodal Graph-based Event Detection and Summarization in Social Media Streams. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 189–192.

[33] Karen Simonyan and Andrew Zisserman. 2014. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 568–576.

[34] C. G. M. Snoek, K. E. A. van de Sande, D. Fontijne, A. Habibian, M. Jain, S. Kordumova, Z. Li, M. Mazloom, S. L. Pintea, R. Tao, D. C. Koelma, and A. W. M. Smeulders. 2013. MediaMill at TRECVID 2013: Searching Concepts, Objects, Instances and Events in Video. In *TRECVID Workshop*.

[35] S. Sunderrajan and B. S. Manjunath. 2016. Context-Aware Hypergraph Modeling for Re-identification and Summarization. *IEEE Transactions on Multimedia* 18, 1 (Jan 2016), 51–63.

[36] Jiajun Wang, Yu-Gang Jiang, Qiang Wang, Kuiyuan Yang, and Chong-Wah Ngo. 2014. Organizing Video Search Results to Adapted Semantic Hierarchies for Topic-based Browsing. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 845–848.

[37] T. Yao, C. W. Ngo, and T. Mei. 2013. Circular Reranking for Visual Search. *IEEE Transactions on Image Processing* 22, 4 (April 2013), 1644–1655.

[38] Soh Yoshida, Takahiro Ogawa, and Miki Haseyama. 2015. Heterogeneous Graph-based Video Search Reranking Using Web Knowledge via Social Media Network. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 871–874.

[39] Y. Zheng, Y. J. Zhang, and H. Larochelle. 2016. A Deep and Autoregressive Approach for Topic Modeling of Multimodal Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 6 (June 2016), 1056–1069.

[40] Lei Zhu, Jialie Shen, and Liang Xie. 2015. Topic Hypergraph Hashing for Mobile Image Retrieval. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 843–846.