LEARNING SEMANTIC-PRESERVING SPACE USING USER PROFILE AND MULTIMODAL MEDIA CONTENT FROM POLITICAL SOCIAL NETWORK

Wei-Hao Chang, Jeng-Lin Li, Chi-Chun Lee

Department of Electrical Engineering, National Tsing Hua University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

ABSTRACT

The use of social media in politics has dramatically changed the way campaigns are run and how elected officials interact with their constituents. An advanced algorithm is required to analyze and understand this large amount of heterogeneous social media data to investigate several key issues, such as stance and strategy, in political science. Most of previous works concentrate their studies using text-as-data approach, where the rich yet heterogeneous information in the user profile, social relationship, and multimodal media content is largely ignored. In this work, we propose a two-branch network that jointly maps the post contents and politician profile into the same latent space, which is trained using a large-margin objective that combines a cross-instance distance constraint with a within-instance semantic-preserving constraint. Our proposed political embedding space can be utilized not only in reliably identifying political spectrum and message type but also in providing a political representation space for interpretable ease-of-visualization.

Index Terms— semantic-preserving space, multimodal media data, social media, politics, large-margin objective

1. INTRODUCTION

Social media plays a pivotal role in electoral campaigns [1, 2]. Research has shown that social media, such as Twitter, has changed drastically the way campaigns are run due to its ability to rapidly disseminate information for politicians to interact with their followers and communities [3]. In fact, political nominees often carefully encode their intended campaign messages on the social media platform to the electorate in order to maximize their influences [4, 5]. Analyzing these message contents to quantitatively understand political subject's stance, value, and even campaign strategy becomes critical in the current era of study in political science.

Computational advancement has further enabled automated analyses of social media content in the political space, specifically the surging use of network learning approach for systematic characterization of social media using neural networks [6]. For the posted media content, the most widely studied data source is the textual content. For example, Wilkerson et al. investigated large scale text data quantitatively and demonstrate that their methodology is capable of analyzing each political party's strategy [7]. Matthew et al. employed unsupervised learning techniques using text-as-data to investigate various issues in political science and demonstrated that pre-processing is key when performing such a study [8]. However, this large body of work primarily focused on structured textual contents ignoring the fact that feeds on social media contain additional data of multimodal types, e.g., rich information in visual content, social context, and social relations. This is partially due to the technical difficulties in capturing the hidden relationship jointly from the heterogeneous types of social media feeds.

Modeling the hidden semantic relationships are critical in bridging the social content to the users interests. Several recent works have worked on integrating heterogeneous data resulting from the social media feeds to advance user interests-social content modeling. For example, Canonical Correlation Analysis (CCA) approach has been adopted to process multimodal data source [9] though this linear approach has been shown to be inadequate to process large amount of data [10]. Other research have developed a variety of frameworks attempting to jointly process these multimodal data, but most perform *late* integration, where each media modality is processed separately at the beginning [11, 12, 13, 14]. In this paper, we present an framework in learning semantic-preserving space for the purpose of jointly capturing latent semantic relationship between user profiles, social links and multimedia contents from political postings.

Specifically, we design a two-branch semantic-preserving network, where one of them represents personal profiles and social links, and another one represents the posted multimodal media content. The network is jointly optimized across these heterogeneous data using a large-margin objective, i.e., a cross-instance distance constraint combining with a within-instance semantic preserving constraint. The learned space simultaneously modeled information of social media feeds (i.e. textual contents and visual contents) and user social contexts. We apply the learned embedding to analyze political messages on Twitter. Our embedding not only obtains promising discriminative power in identifying message types and political spectrum, but also offers a visualization of joint representation for political subjects on social media.

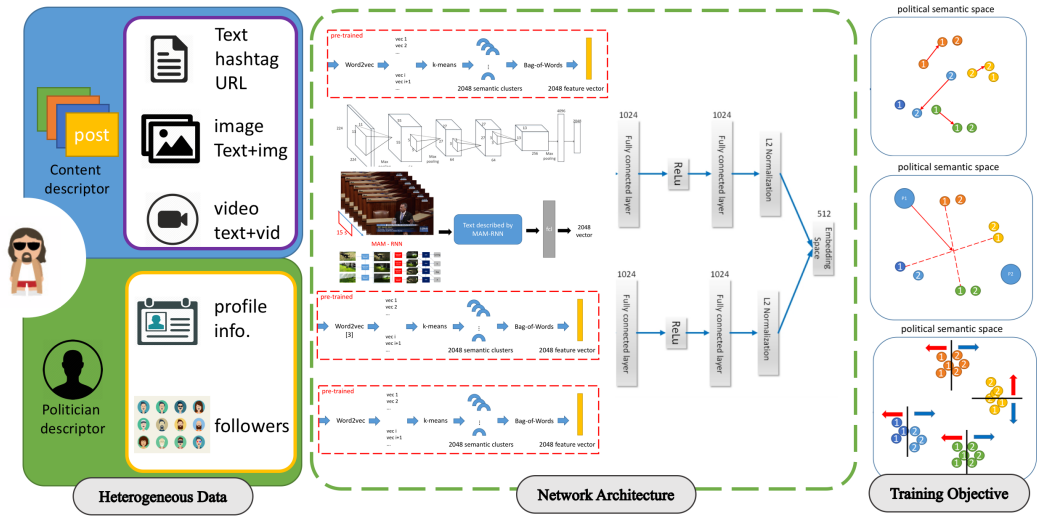

Fig. 1. The illustration of our proposed political semantic space network.

Fig. 2. An example of social media posts.

2. RESEARCH METHODOLOGY

In this section, we present the details in learning our semantic-preserving space. Our dataset is introduced in Section 2.1. Second, we illustrate the network structure in Section 2.2. The learning objective is presented in Section 2.3.

2.1. Dataset: Twitters Political Posts

We use the Crowdflower's Data For Everyone Library as the political dataset in this work. It provides 5000 text messages from 56 unique politician's social media accounts. Each message has been annotated with its purpose, partisanship, and intended audience. The message type are categorized into nine groups: attack, constituency, information, media, mobilization, personal, policy, support, other.

In order to learn a multimodal semantic political space for these 56 politicians, we further retrive additional data source from these 5000 Twits to include heterogeneous types of data (see Fig 2). Each message contains two types of data. The first is the message content, and the second one is political profile. The content of each post can be further categorized as extrinsic or intrinsic data. Intrinsic data indicates the actual content of the post, e.g., text, hashtag, image, or video; one thing to note that hashtag is used due to past works in indicating its inclusion of more semantic ties to the post compared to the textual content [15]. Extrinsic content data is information such as the content retrieved from the embedded shortened external links within the post. In terms of politician descriptors, except for conventional user's personal profile (user tags, gender, location, and etc), we also take into account of the user's social relationship, i.e., all of their follower's account. This consititues our entire dataset.

2.2. Content and Profile Descriptors

The architecture of our proposed network in learning the semantic-preserving space is shown in the middle part of Fig 1. The network includes three branches (text, images, video) in the content descriptor and two branches (user profile, social links) in the politician descriptor. We first describe the encoder network for each our post modality. In general, we first train a Word2Vec (skip-gram, dim=2048, window_size=8, neg_samples=100) with K-means bag-of-word to generate a 2048-dimension text encoder from our dataset.

The textual content, which includes text content, hashtag, and URL (the main article that URL goes to) module of the content is, hence, encoded each using this 2048-dimensional text encoder. The images are encoded using a convolutional neural network (CNN). We design a ConvNet consists of 5 convolution layers, which takes in 224x224 raw images and encode it 2048 dimensional feature vector. We use the MAM-RNN as video semantic encoder [16] ,which is a video captioning RNN model using multi-level attention scheme that transcribes the video to captions. The video caption's transcripts are fed into the Word2Vec encoder to generate the video embedding. All of these encoded descriptors are integrated by feeding into a two fully-connected layers to generate 1024 dimensional content-descriptors. Furthermore, gen-

Table 1. Performance using the learned multimodal embedding space for message type classifications.

Components	Performance: UAR (9 class)									
text	Poli.	Pers.	Supp.	Info.	Medi.	Attk.	Cnsti.	Mobil.	Other	Average
B:post+user	0.408	0.308	0.420	0.312	0.320	0.457	0.442	0.347	0.314	0.37
B+hashtag	0.417	0.300	0.424	0.304	0.310	0.485	0.447	0.345	0.298	0.37
B+URL	0.431	0.319	0.456	0.315	0.342	0.489	0.474	0.374	0.306	0.39
Hybrid1 (textual)	0.494	0.320	0.477	0.320	0.356	0.532	0.497	0.374	0.316	**0.41**
image	Poli.	Pers.	Supp.	Info.	Medi.	Attk.	Cnsti.	Mobil.	Other	Average
image	0.392	0.284	0.403	0.285	0.295	0.428	0.436	0.342	0.283	0.35
B+image	0.448	0.303	0.430	0.288	0.301	0.462	0.465	0.349	0.284	0.37
Hybrid2 (Hybrid1+vision)	0.509	0.325	0.511	0.317	0.352	0.535	0.528	0.393	0.309	**0.42**
video	Poli.	Pers.	Supp.	Info.	Medi.	Attk.	Cnsti.	Mobil.	Other	Average
MAM-RNN	0.436	0.294	0.418	0.279	0.292	0.449	0.452	0.339	0.276	0.36
Hybrid3 (Hybrid1+MAM-RNN)	0.421	0.304	0.433	0.316	0.323	0.505	0.503	0.379	0.321	**0.39**
Hybrid (1+2+3)	0.521	0.333	0.523	0.325	0.361	0.547	0.541	0.402	0.316	**0.43**

erating the politician descriptor is straightforward. We use the same Word2Vec with K-means bag-of-word encoder approach to generate a 2048-dimension feature vector on politician name and their social links, then two fully connected layers are followed to output a 1024-dimensional feature vector as the politician descriptor.

In summary, for every post, we derive two types of descriptors: content descriptors and profile descriptors. The content descriptor includes text, image and video embedded vectors, that is learned in parallel and integrated through fully-connected dense layers; the profile descriptors include user name and social links, which are integrated also via dense layers. These two branch of descriptors are fed into another fully-connected layer to derive the final *political semantic-preserving space*, the training objective is detailed in the next section.

2.3. Network Architecture: Training Objective

Our training objective is based on a stochastic hinge loss includes an cross-instance distance constraint combined with a within-instance semantic-preserving constraint. The derivation of the semantic-preserving space is based on the assumption that he space should position individual politician's posts to be close to each other, at the same time, the *semantically-same* type of messages should be close to each other. The illustrative diagram is shown in the right part of Fig 1.

2.3.1. Cross-instance Distance Constraint

Within the post space, P_i, given an politician u_i, his/her posting, termed as positive samples, p_j^+, and others as negative samples, p_k. We want the distance between u_i and each positive sample post p_j^+ to be smaller than the distance between u_i and each negative sample p_k by an enforced margin m:

$$d(u_i, p_j^+) + m < d(u_i, p_k) \forall p_j^+ \in P_i^+, \forall p_k^- \in P_i^- \quad (1)$$

$d()$ is the Euclidean distance between 2 vectors.

2.3.2. Within-instance Semantic-preserving Constraint

We add additional constraints that same message type posts should be closer to each other than to any other types of posts. Further, within the same type of messages, the same political spectrum should have posts tend to be closer to each other within this space. We add this semantic-preserving constraint to preserve the within-instance structure. Let S represents a set of semantic clusters, and N_{pi}^s denotes a set of neighborhood posts of p_i that belong to the same semantic cluster s. Semantic clusters are the message types and the political spectrum for a given message. Then, we similarly enforce a margin of m between N_{pi}^s and any point outside of s:

$$d(p_i, p_j) + m < d(p_i, p_k) \forall p_j \in N\{P_i\}^s \quad (2)$$

Hence, we combine these two constraints into our training objective, the complete loss function to update the fully-connected layers and the final embedding space is given by:

$$\begin{aligned} L(U, P) = &\sum_{i,j,k} max[0, m + d(u_i, p_j^+) - d(u_i, p_k^-)] \\ &+ \lambda \sum_{i,j,k} max[0, m + d(p_i, p_j) - d(p_i, p_k)] \end{aligned} \quad (3)$$

where λ controls the importance of the two loss terms.

3. EXPERIMENTAL RESULT AND ANALYSIS

3.1. Experimental Settings

Network Settings. ConvNet is trained on cropped images of size 224x224 with 4 times data augmentation (GaussianNoise, RandomHorizontalFlip/VerticalFlip, Normalize). The loss function consists a triplet terms: cross-instance politician constraint, within-instance message type constraint, and within-instance political party constraint. We sample 64 triplets within each minibatch. SGD is used to train the whole network, and the learning rate is 0.01, the weight decay is 1e-5 and momentum is 0.9. We apply batch normalization and dropout layer with 50% in our network and set enforced margin m to 1e+4 as the max tolerable distance.

Table 2. Political spectrum classification result.

Message type	Performance: UAR (2 class)		
	SVM	DNN	Our method
Policy	66.33	71.54	**82.43**
Personal	50.23	51.34	**53.33**
Support	68.22	72.72	**79.53**
Information	50.78	52.12	**53.33**
Media	52.11	53.23	**59.41**
Attack	74.31	83.12	**88.67**
Constituency	71.89	76.64	**82.85**
Mobilization	56.44	59.23	**62.41**
Other	52.12	51.11	**52.81**
Average	52.28	63.45	**68.31**

We conduct two different experiments. The first is to classify the message types using different multimodal embedding, and the second is to classify and visualize the political spectrum of the subjects in our dataset using the learned space. We use 5-fold cross validation use SVM (kernel=linear, c=1) as our final classifier.

3.2. Experiment 1: Message Types Classification

Table 1 summarizes our nine-class message types classification (change level = 11.11%) using various branches of embedding (Fig 1) of our framework. Overall, our proposed multimodal (inclusion of intrinsic extrinsic multimodal post content data with user profile) obtains the best classification rates of 43%. The baseline (B) here represents the most widely-used method, i.e., post textual content and user profile. Our method is 6% over the baseline method. We also observe that inclusion of extrinsic data (i.e., the content retrieved from the linked URL) provides important complementary information in decoding the types of the messages in this context.

While Chen et al. has claimed that visual content may be insufficient in interpreting social media content [17], we observe that in our experiment, by including image representation improve slightly beyond using textual components only (42% versus 37%). However, the diverse images used in the political space is much more complex requiring future works in better handle encoding the raw image data. In terms of video, MAM-RNN with textual components achieve 39% recognition rates over baseline 37%. Finally, combining all captures the largest amount of relevant information, which reflects in the improvement of message types classifications.

3.3. Experiment 2: Political Party Classification

Table 2 summarizes the political party classification results. The baseline uses concatenation of content vector and profile descriptors as input and takes SVM as 2-class classifier, and DNN indicates using 3 fully connected layers (1024, 512, 2) on the same descriptor input. λ sets 1e+3 to emphasize post information. Our proposed method takes the final embedding vector to feed into the SVM. Our proposed semantic-preseving space achieves the best accuracy of 68.31% on average over nine different message types for political spectrum classification. We observe a varying degree of discrminability between different message types. For example, message types such as "policy", "attach", and "constituency" obtains high UAR. The use of our framework in improving the discriminative representation of social posts between the two parties are also apparent for the message type of "policy", "media", and "mobilization".

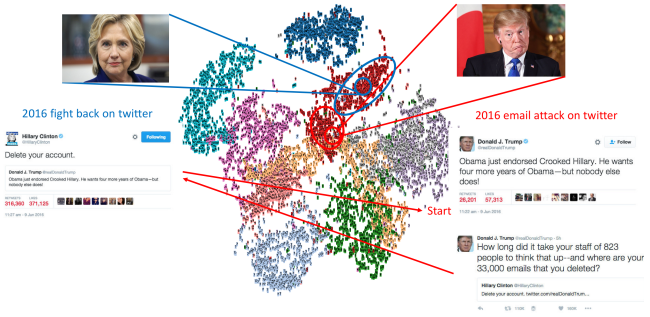

Fig. 3. Visualization of political semantic space. Red points represent for "attack" and our method can separate the posts from 2 political parties on 2016 U.S. election.

Our embedding is not only useful for a variety of classification tasks. We can further plot the multimodal nature of the social network feeds joint together to intuitively visualize the relationship between politicians and their posts. As an example shows in Figure 3, red points represent attack message and further cluster into 2 political party, which shows that our model can separate the email attack on Twitter from 2 political parties in the 2016 U.S. election.

4. CONCLUSIONS

In this paper, we propose a novel learning framework to derive semantic-preserving multimodal political social media space. The learned framework effectively captures the hidden semantic relationship between social media contents and user profiles. We use a two-branch neural network learned with a large-margin objective function (cross-instance constraint and within-instance constraint) to train the multimodal data source (textual content, visual content: images and videos, and social relation) in an end-to-end manner. The jointly-learned space not only obtains improvement in message type classification and political spectrum recognition, but also provides an intuitive visualization framework that has already embedded all relevant yet heterogeneous information.

As for future work, we will further analyze the longitudinal aspect of social media posting of an election to further the strategy used in the campaign and to identify potential trend and public opinions in order to provide a reliable meter to evaluate and predict the elected politician.

5. REFERENCES

[1] Megan Fountain, *Social Media and its Effects in Politics: The Factors that Influence Social Media use for Political News and Social Media use Influencing Political Participation*, Ph.D. thesis, The Ohio State University, 2017.

[2] Joseph Kahne and Benjamin Bowyer, "The political significance of social media activity and social networks," *Political Communication*, pp. 1–24, 2018.

[3] Bente Kalsnes, Anders Olof Larsson, and Gunn Enli, "The social media logic of political interaction: Exploring citizens and politicians relationship on facebook and twitter," *First Monday*, vol. 22, no. 2, 2017.

[4] Daniel Halpern, Sebastián Valenzuela, and James E Katz, "We face, i tweet: How different social media influence political participation through collective and internal efficacy," *Journal of Computer-Mediated Communication*, vol. 22, no. 6, pp. 320–336, 2017.

[5] Kaitlin Vonderschmitt, "The growing use of social media in political campaigns: How to use facebook, twitter and youtube to create an effective social media campaign," 2012.

[6] Ibrahim Uysal and W Bruce Croft, "User oriented tweet ranking: a filtering approach to microblogs," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 2261–2264.

[7] John Wilkerson and Andreu Casas, "Large-scale computerized text analysis in political science: Opportunities and challenges," *Annual Review of Political Science*, vol. 20, pp. 529–544, 2017.

[8] Matthew J Denny and Arthur Spirling, "Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it," *Political Analysis*, vol. 26, no. 2, pp. 168–189, 2018.

[9] Benjamin Klein, Guy Lev, Gil Sadeh, and Lior Wolf, "Fisher vectors derived from hybrid gaussian-laplacian mixture models for image annotation," *arXiv preprint arXiv:1411.7399*, 2014.

[10] Zhuang Ma, Yichao Lu, and Dean Foster, "Finding linear structure in large datasets with scalable canonical correlation analysis," in *International Conference on Machine Learning*, 2015, pp. 169–178.

[11] Yunchao Gong, Qifa Ke, Michael Isard, and Svetlana Lazebnik, "A multi-view embedding space for modeling internet images, tags, and their semantics," *International journal of computer vision*, vol. 106, no. 2, pp. 210–233, 2014.

[12] Tao Chen, Xiangnan He, and Min-Yen Kan, "Context-aware image tweet modelling and recommendation," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 1018–1027.

[13] Chao Wu, Jia Jia, Wenwu Zhu, Xu Chen, Bowen Yang, and Yaoxue Zhang, "Affective contextual mobile recommender system," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 1375–1384.

[14] Yunchao Gong, Liwei Wang, Micah Hodosh, Julia Hockenmaier, and Svetlana Lazebnik, "Improving image-sentence embeddings using large weakly annotated photo collections," in *European Conference on Computer Vision*. Springer, 2014, pp. 529–545.

[15] David Laniado and Peter Mika, "Making sense of twitter," in *International Semantic Web Conference*. Springer, 2010, pp. 470–485.

[16] Xuelong Li, Bin Zhao, and Xiaoqiang Lu, "Mam-rnn: multi-level attention model based rnn for video captioning," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.

[17] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, and Xian-Sheng Hua, "Deep ctr prediction in display advertising," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 811–820.