ICMR '20, October 26–29, 2020, Dublin, Ireland
Special Session 4: Knowledge-Driven Analysis and Retrieval on Multimedia
Proceedings published June 8, 2020

Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks

Youze Wang[1], Shengsheng Qian[2,3], Jun Hu[1], Quan Fang[2,3], Changsheng Xu[1,2,3,4]
[1]Hefei University of Technology
[2]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
[4]Peng Cheng Laboratory, ShenZhen, China
youze.wang@mail.hfut.edu.cn,shengsheng.qian@nlpr.ia.ac.cn
hujunxianligong@gmail.com,{qfang,csxu}@nlpr.ia.ac.cn

ICMR'20 (ACM International Conference on Multimedia Retrieval)

220428 Chia-Chun Ho

Outline

Introduction

Related Works

Problem Statement

Methodology

Experiments

Conclusion

Comments

Introduction

Fake news intro

- Social media websites have fostered various fake news which usually contain misrepresented or even forged multimedia content.
  - Mislead the reader and get rapid spread.
  - Mislead public opinion to damage the credibility of the government on purpose.
- It's necessary and urgent to use an automatic detector to prevent fake news from causing serious negative effects and make users receive truthful information.

Introduction
Challenges

- Most of the existing deep learning methods only capture local semantic features in small sliding windows (short messages or word-level syntactics).
    - Ignore the structural information of posts which is a very important aspect for fake news detection.
- Some posts may have many words and understanding the semantics of them need to model non-consecutive phrases and long-range word dependency.

Introduction
Challenges

- Fake news detection needs to detect fake news from various field which the model maybe seen.
- Existing methods typically focus on inferring clues from the post text content.
- Think little of the visual information and background knowledge of posts which humans also use in judging the credibility of an event.
- How to acquire the background knowledge of the post text content, and fuse the textual information, knowledge concepts and visual information of the post in a principled way is the key for fake news detection.

Introduction
Example

- Keywords: "Michael Bloomberg", “impeachment", "Donald Trump"
    - Indicates that the semantics may be related to "Michael Bloomberg supported the impeachment to Donald Trump".
- These keywords are not grouped together and distributed throughout the whole post.
- Hard to capture the dependency of semantic and structure information in a small sliding window.

Figure 1: An example of multimodal tweets. We can obtain a lot of information, such as Michael Bloomberg, Democrat, impeachment, Donald Trump from the textual content, and see a speaker, the Stars and the Stripes as visual information, and think of the speaker is a politician, Donald Trump is American President from the background knowledge.

Introduction
KMGCN

- Propose a Knowledge-driven Multimodal Graph Convolutional Network (KMGCN).
  - Modeling posts as graphs data structure, and combining the textual information, knowledge concepts and visual information into a unified deep model.
- To capture long-range semantic features for better content representations:
  - Model each post content as graph rather than word sequences.
- To make full use of the background knowledge & multimodal information:
  - Utilize object detection techniques to extract object in image as visual words and obtain knowledge concepts through knowledge distillation.

Introduction

Contributions

- Propose an end-to-end Knowledge-driven Multimodal Graph Convolutional Network.
  - To model the semantic-level representations by jointly modeling textual information, knowledge concepts and visual information into a unified deep model.
- Model multimodal posts as graphs in classification tasks and propose a multimodal GCN.
  - Capture non-consecutive and long-range semantic relations.
  - Knowledge distillation is used to provide supplementary knowledge concept which can generalize well for the newly emerged posts.

Related Works
Fake News Detection

Problem Statement

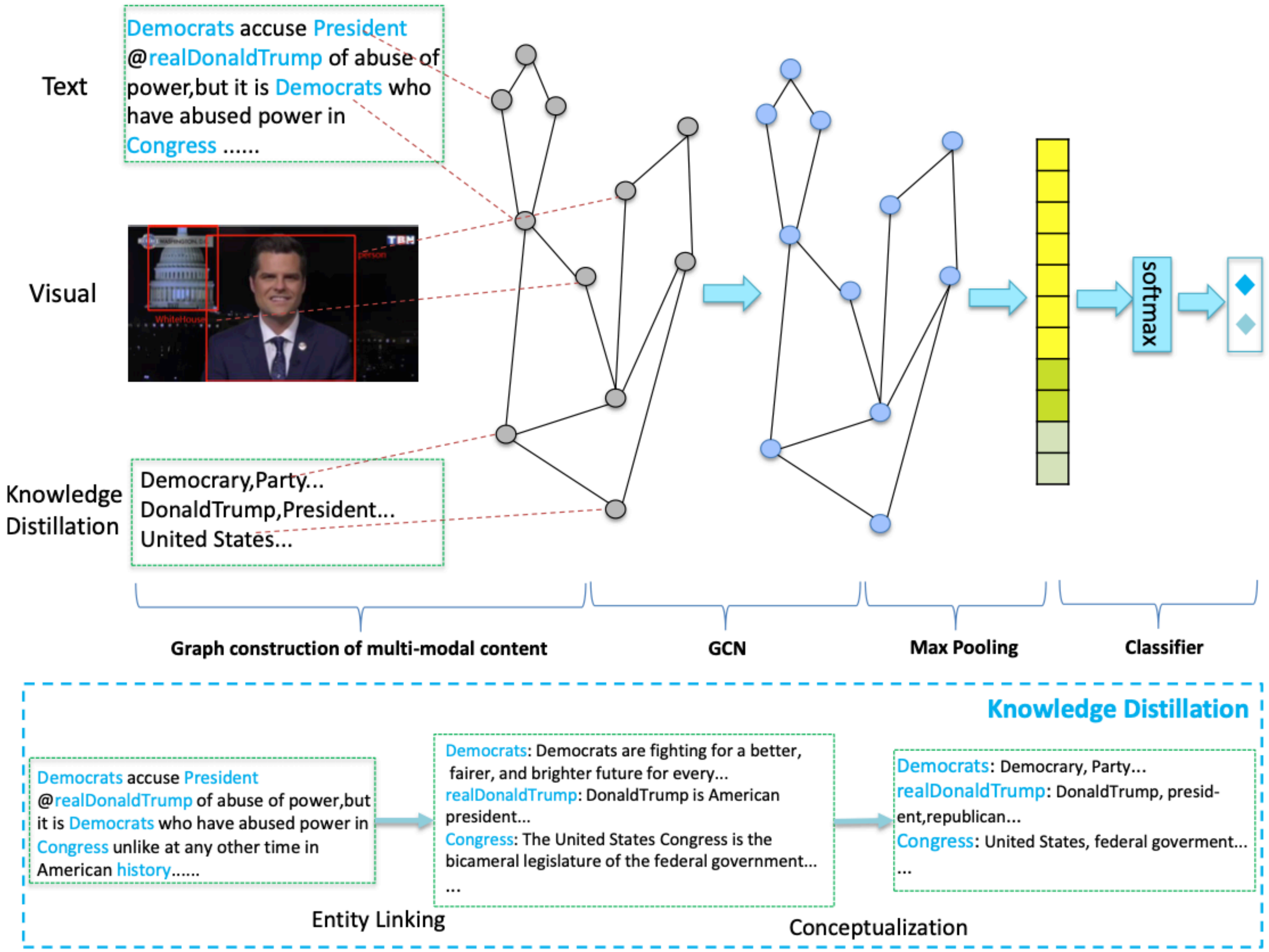
Problem

- Defined as a binary classification problem.
  - The goal of model is to identify whether a post is fake or not at the post-level.
- Given a set of multimedia posts from social media $D = \{p_1, \ldots, p_N\}$.
  - p_i is a post which consists of a set of words and corresponding visual information.
  - N represents the number of posts.
- Learn a model $f : \mathscr{D} \rightarrow \mathscr{Y}$, to classify each post into the pre-defined categories $\mathscr{Y} = \{0,1\}$. (0: real, 1: fake)

Methodology

Knowledge-driven Multimodal Graph Convolutional Networks (KMGCN)

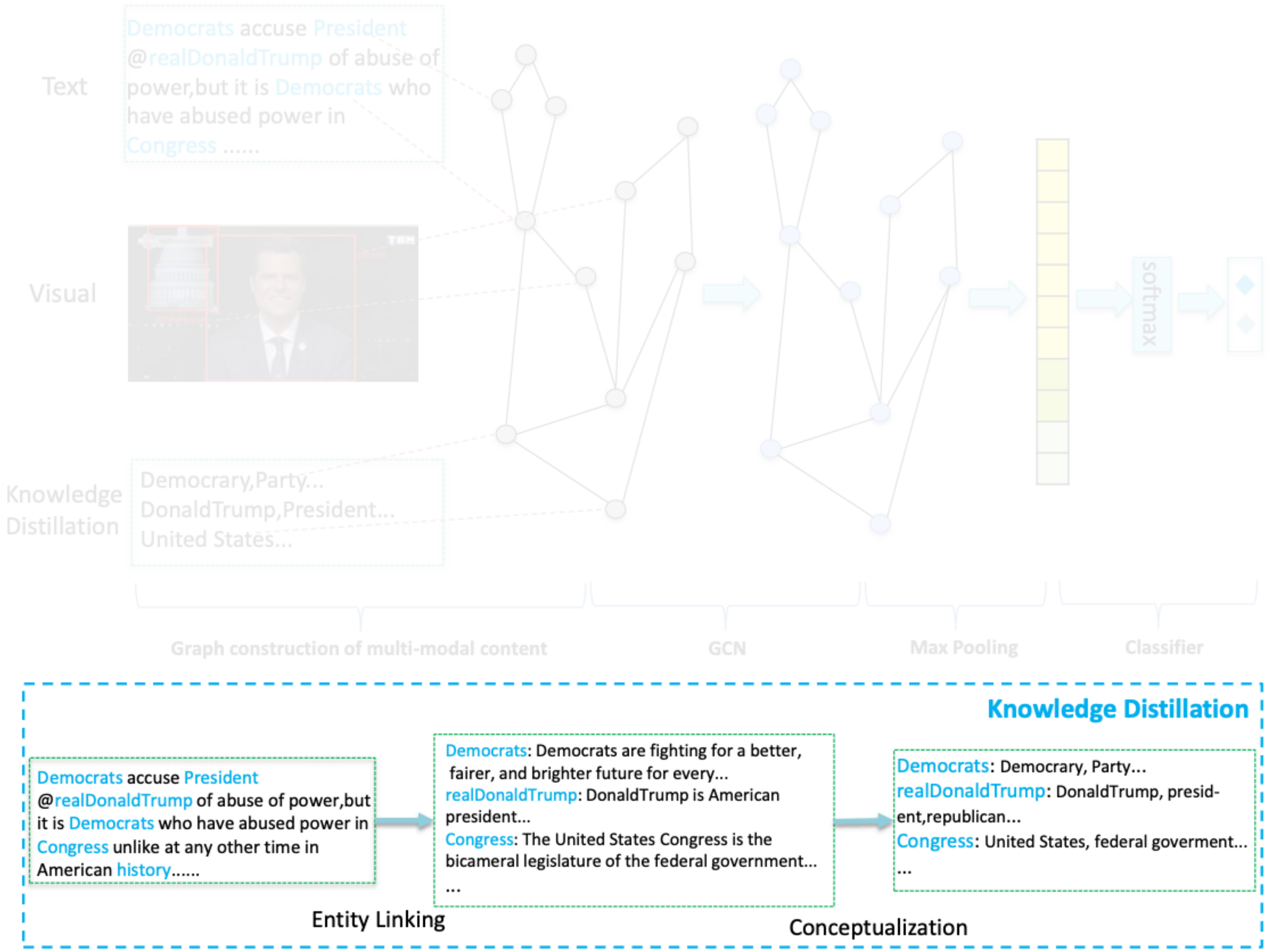
- Knowledge Distillation
- Graph Construction of Multimodal Content
- Knowledge-driven Multimodal Graph Convolution Network



Methodology

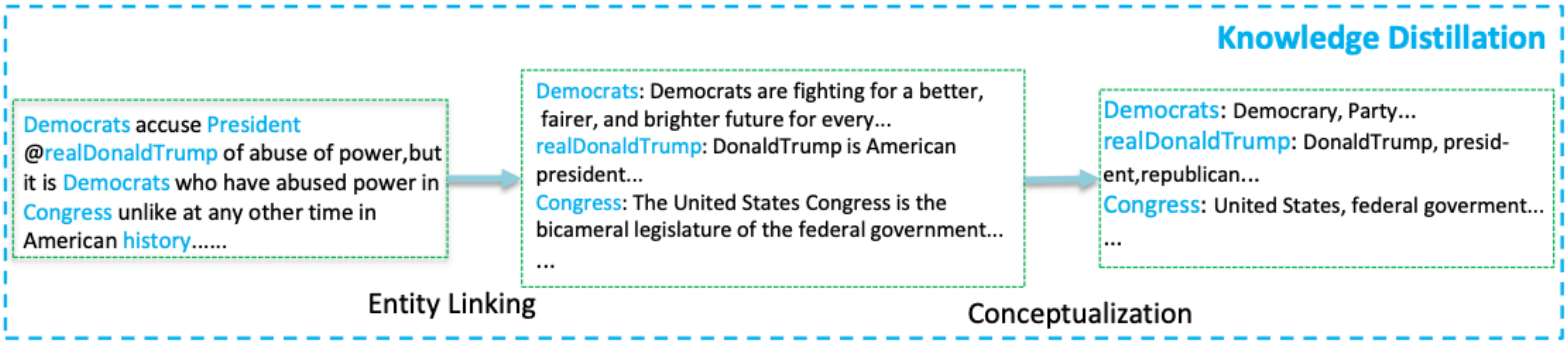
Knowledge-driven Multimodal Graph Convolutional Networks (KMGCN)

- Knowledge Distillation
- Graph Construction of Multimodal Content
- Knowledge-driven Multimodal Graph Convolution Network



Methodology
Knowledge Distillation

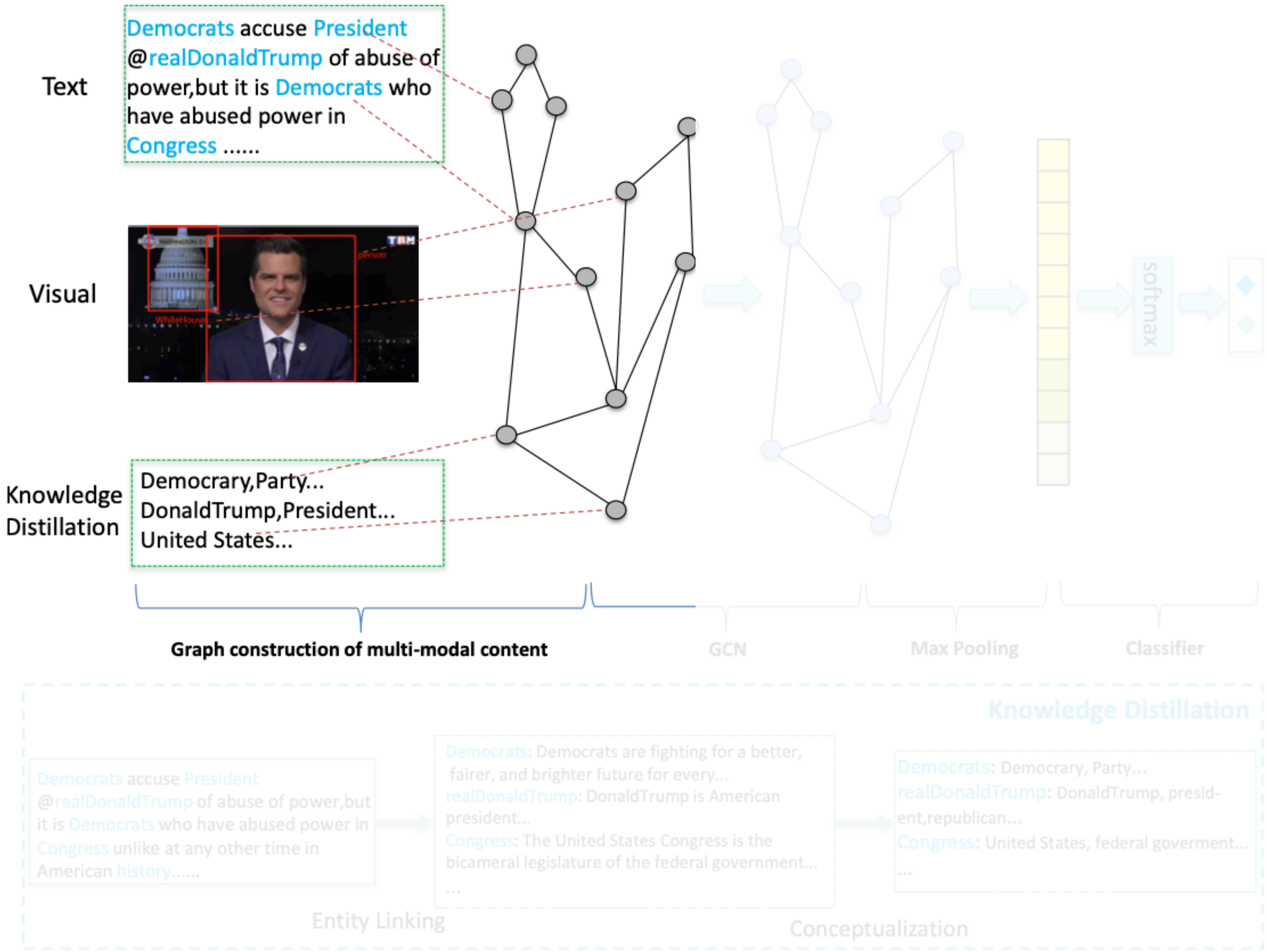
- Given a post text, hope to find a concept set C relevant to it.
- Process of knowledge distillation as follow.



Methodology

Knowledge-driven Multimodal Graph Convolutional Networks (KMGCN)

- Knowledge Distillation
- Graph Construction of Multimodal Content
- Knowledge-driven Multimodal Graph Convolution Network



Methodology

Graph Construction of Multimodal Content

- Given a post, the words are taken as the graph nodes and the relationship between words is taken as edges.
- Employ point-wise mutual information (PMI) to calculate the weights of edges, which can preserve the global word co-occurrence information.
- In detail, employ a fixed-size window on all posts content for gathering word co-occurrence statistics.

Methodology
Graph Construction of Multimodal Content

- Calculate the PMI of word pairs as follows:
  - $p(w_i) = \dfrac{W(w_i)}{|W|}, p(w_i, w_j) = \dfrac{W(w_i, w_j)}{|W|}, PMI(w_i, w_j) = \log \dfrac{p(w_i, w_j)}{p(w_i)p(w_j)}$
  - $W(w_i)$: # of sliding windows that contain the word w_i
  - $W(w_i, w_j)$: # of sliding windows that contain both the word w_i and w_j
  - $|W|$: total # of sliding windows.

Methodology

Graph Construction of Multimodal Content

- Note that the statistics are based on the global corpus rather than a specific post content.

- PMI scores can reflect the correlation between words, and positive PMI scores implies high semantic correlations.

- Therefore, only preserve edges with positive PMI scores and discard those with non-positive PMI scores:

- $$A_{ij} = \begin{cases} PMI(w_i, w_j) & PMI(w_i, w_j) > 0 \\ 0 & PMI(w_i, w_j) \leq 0 \end{cases}$$

Methodology

Graph Construction of Multimodal Content

- In addition to the text content, also utilize the visual content to facilitate the fake news detection model by extracting semantic objects in images as visual words.
- Employ YOLO-v3 detector to detect several semantic objects in each image.
  - The labels of detected objects, such as "person" and "gun", are treated as words that occurred in a post and are added into the text content of each post.

Methodology

Graph Construction of Multimodal Content

- Besides textual / visual information, also obtain a set of knowledge concepts from knowledge graphs for every entity in a post to complement the semantic representation of the post.

- Like visual words, also add knowledge concepts into the text content of each post.

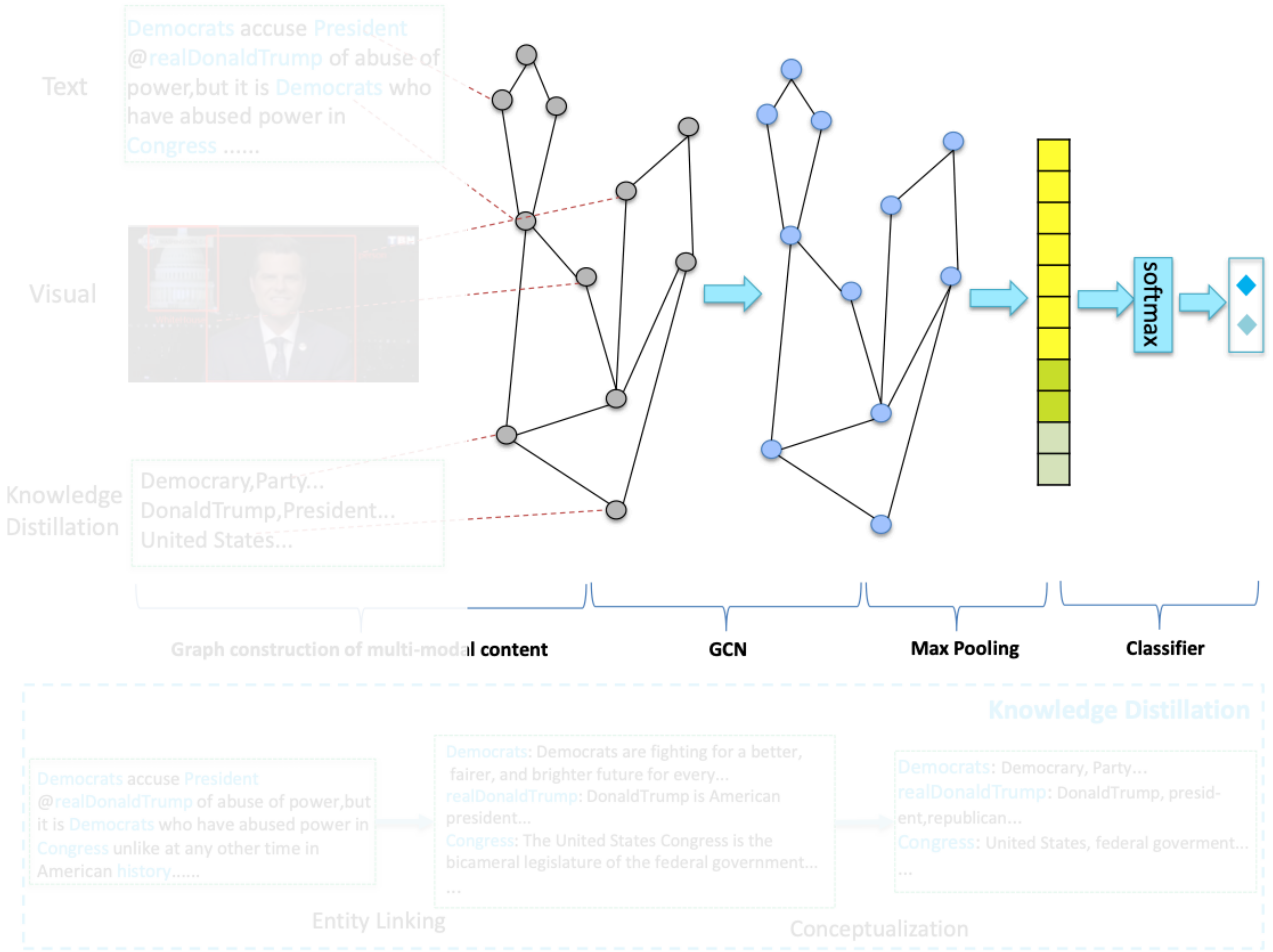
- In the end, build a graph which contains textual words, visual words and knowledge concepts by PMI for each post.

- After this process, obtain an adjacency matrix A for each post.

Methodology

Knowledge-driven Multimodal Graph Convolutional Networks (KMGCN)

- Knowledge Distillation
- Graph Construction of Multimodal Content
- Knowledge-driven Multimodal Graph Convolution Network



Methodology

Knowledge-driven Multimodal Graph Convolution Network

- Consider an undirected graph $G = (V, E)$
- V are sets of nodes and E are sets of edges.
- Every node assumed to be connected to itself, i.e., $(v, v) \in \mathbb{E}$ for any v.
- Each node is associated with a d-dimensional feature vector and use a feature matrix X to represent the feature of all vertices, each row correspond to one node.

Methodology

Knowledge-driven Multimodal Graph Convolution Network

- Adjacency matrix A to indicate the edge set E.
  - A_{ij} is the weight of the edge between the v_i and v_j.
- Degree matrix D is a diagonal matrix, $D_{ij} = \sum_j A_{ij}$.
- Based on A & D, each GCN layer input feature matrix $Z^{(j)}$ & output a higher-order feature matrix $Z^{(j+1)}$ for vertices as follows:
  - $Z^{(0)} = X$, $Z^{(j+1)} = \sigma(D^{-\frac{1}{2}}(I + A)D^{-\frac{1}{2}}Z^{(j)}W)$
  - For X for each post, employ the distributed Word2Vec representation for words.

Methodology

Knowledge-driven Multimodal Graph Convolution Network

- After two layers of GCN, choose a global mean pooling to aggregate the vertices of each graph and get the representation vector O of posts.

- Finally, feed each post p_i's representation vector O_i into a binary classifier and get prediction:

  - $r_i = \mathscr{F}(O_i)$

- To calculate the classification loss employ cross-entropy as follow:

  - $\mathscr{L} = \sum_{\mathrm{i}=1}^{N} -[y_{\mathrm{i}} \times \log(r_{\mathrm{i}}) + (1 - y_{\mathrm{i}}) \times \log(1 - r_{\mathrm{i}})]$

Experiments

Datasets

News	PHEME	WEIBO
Fake News	1972	2313
Real News	3830	2351
Images	3670	3989

- PHEME
  - collected based on 5 breaking news, and news contain a set of claims.
- Weibo
  - collected based on the claims reported on Weibo, where each claim contains text, image URL, response and so on.
- Train : Test = 7:3

Experiments

Baselines

- SVM-TS: proposes the linear SVM classifier that uses time-series structures to model the variation of social context features.
- GRU: uses a multilayer generic GRU network to model the microblog as a variable-length time series, which is effective for the early detection of rumors.
- CNN: uses a convolution network to learn rumor representations by framing the relevant posts as fixed-length sequence.
- TextGCN: use GCN to classify documents. The whole corpus is modeled as a heterogeneous graph. It combines graph neural network to learn words and document embedding.
- EANN: utilizing an adversarial method to remove event-specific features from post representation.

Experiments

Quantitative Evaluation

Dataset	Methods	Accuracy	Precision	Recall	F1
WEIBO	SVM-TS	0.6312	0.6329	0.6301	0.6309
	CNN	0.7112	0.713	0.7112	0.711
	EANN	0.7212	0.7353	0.7228	0.7160
	GRU	0.7927	0.8139	0.7927	0.7891
	TextGCN	0.8571	0.8634	0.8576	0.8565
	KMGCN	**0.8863**	**0.9100**	**0.9645**	**0.8834**
PHEME	SVM-TS	0.6399	0.6391	0.6211	0.6395
	CNN	0.7007	0.7413	0.7074	0.6896
	EANN	0.7177	0.7382	0.7179	0.7104
	TextGCN	0.8282	0.8274	0.8283	0.8277
	GRU	0.8374	0.8382	0.8374	0.8312
	KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

- SVM-TS model performs worst among all methods, which is possible that the hand-crafted features are weak and not enough to identify fake news.

- CNN is only better than SVM-TS on two datasets, which is probably because the CNN ignores the long-range semantic relations among words and local feature is not enough to make judgment for a post.

Experiments

Quantitative Evaluation

Dataset	Methods	Accuracy	Precision	Recall	F1
WEIBO	SVM-TS	0.6312	0.6329	0.6301	0.6309
	CNN	0.7112	0.713	0.7112	0.711
	EANN	0.7212	0.7353	0.7228	0.7160
	GRU	0.7927	0.8139	0.7927	0.7891
	TextGCN	0.8571	0.8634	0.8576	0.8565
	KMGCN	**0.8863**	**0.9100**	**0.9645**	**0.8834**
PHEME	SVM-TS	0.6399	0.6391	0.6211	0.6395
	CNN	0.7007	0.7413	0.7074	0.6896
	EANN	0.7177	0.7382	0.7179	0.7104
	TextGCN	0.8282	0.8274	0.8283	0.8277
	GRU	0.8374	0.8382	0.8374	0.8312
	KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

- EANN > CNN, because EANN employs TextCNN to extract textual feature and VGG-19 to extract visual feature.
  - Provides complementary information to improve fake news detection.
- TextGCN performs best in all baselines in Weibo and performs better in PHEME.
  - Shows that the graph structure can effectively capture word co-occurrences and document-word relations by the flexible graph convolutional network.

Experiments

Quantitative Evaluation

Dataset	Methods	Accuracy	Precision	Recall	F1
WEIBO	SVM-TS	0.6312	0.6329	0.6301	0.6309
	CNN	0.7112	0.713	0.7112	0.711
	EANN	0.7212	0.7353	0.7228	0.7160
	GRU	0.7927	0.8139	0.7927	0.7891
	TextGCN	0.8571	0.8634	0.8576	0.8565
	KMGCN	**0.8863**	**0.9100**	**0.9645**	**0.8834**
PHEME	SVM-TS	0.6399	0.6391	0.6211	0.6395
	CNN	0.7007	0.7413	0.7074	0.6896
	EANN	0.7177	0.7382	0.7179	0.7104
	TextGCN	0.8282	0.8274	0.8283	0.8277
	GRU	0.8374	0.8382	0.8374	0.8312
	KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

- GRU > CNN on two datasets, because RNN can inherently deal with the variable-length sequence of posts, CNN needs more data to make judgment.
- KMGCN has achieved the best performance compared with all the baselines.
  - KMGCN uses visual information & knowledge concepts to enhance semantic information of post.
  - Multimodal GCNs can better capture non-consecutive phrases and word dependency to obtain more semantic representations.

Experiments

Comparison among KMGCN variants

- -NoKD: removes the knowledge distillation.
- -NoVisual: removes visual information.
- -NoKDVisual: remove both.

Methods	Accuracy	Precision	Recall	F1
textGCN	0.8571	0.8634	0.8576	0.8565
KMGCN-NoKDVisual	0.8713	0.8718	0.974	0.8692
KMGCN-NoVisual	0.8792	0.8748	0.9712	0.8733
KMGCN-NoKD	0.8799	0.8799	0.9327	0.8761
KMGCN	**0.8863**	**0.91**	**0.9645**	**0.8834**

Weibo

Methods	Accuracy	Precision	Recall	F1
GRU	0.8374	0.8382	0.8374	0.8312
KMGCN-NoKDVisual	0.8553	0.8552	0.8089	0.8323
KMGCN-NoVisual	0.8621	0.852	0.8166	0.8339
KMGCN-NoKD	0.8690	0.8675	0.8690	0.8677
KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

PHEME

Experiments

Comparison among KMGCN variants

- Find -NoKDVisual is obviously better than the best of baselines except for Recall on PHEME.
  - Show graph construction as input can effectively capture the long-range semantic relations among words.

Methods	Accuracy	Precision	Recall	F1
textGCN	0.8571	0.8634	0.8576	0.8565
KMGCN-NoKDVisual	0.8713	0.8718	0.974	0.8692
KMGCN-NoVisual	0.8792	0.8748	0.9712	0.8733
KMGCN-NoKD	0.8799	0.8799	0.9327	0.8761
KMGCN	**0.8863**	**0.91**	**0.9645**	**0.8834**

Weibo

Methods	Accuracy	Precision	Recall	F1
GRU	0.8374	0.8382	0.8374	0.8312
KMGCN-NoKDVisual	0.8553	0.8552	0.8089	0.8323
KMGCN-NoVisual	0.8621	0.852	0.8166	0.8339
KMGCN-NoKD	0.8690	0.8675	0.8690	0.8677
KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

PHEME

Experiments

Comparison among KMGCN variants

- Find that the -NoVisual leads to the model's declines in two datasets.
  - Although visual information does not bring huge performance improvement.
    - It can consistently provide complementary information.

Methods	Accuracy	Precision	Recall	F1
textGCN	0.8571	0.8634	0.8576	0.8565
KMGCN-NoKDVisual	0.8713	0.8718	0.974	0.8692
KMGCN-NoVisual	0.8792	0.8748	0.9712	0.8733
KMGCN-NoKD	0.8799	0.8799	0.9327	0.8761
KMGCN	**0.8863**	**0.91**	**0.9645**	**0.8834**

Weibo

Methods	Accuracy	Precision	Recall	F1
GRU	0.8374	0.8382	0.8374	0.8312
KMGCN-NoKDVisual	0.8553	0.8552	0.8089	0.8323
KMGCN-NoVisual	0.8621	0.852	0.8166	0.8339
KMGCN-NoKD	0.8690	0.8675	0.8690	0.8677
KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

PHEME

Experiments

Comparison among KMGCN variants

- Find the KMGCN achieves consistently better results than KMGCN-NoKD in both datasets.
  - Indicates that the knowledge information is an important kind of complementary information.

Methods	Accuracy	Precision	Recall	F1
textGCN	0.8571	0.8634	0.8576	0.8565
KMGCN-NoKDVisual	0.8713	0.8718	0.974	0.8692
KMGCN-NoVisual	0.8792	0.8748	0.9712	0.8733
KMGCN-NoKD	0.8799	0.8799	0.9327	0.8761
KMGCN	**0.8863**	**0.91**	**0.9645**	**0.8834**

Weibo

Methods	Accuracy	Precision	Recall	F1
GRU	0.8374	0.8382	0.8374	0.8312
KMGCN-NoKDVisual	0.8553	0.8552	0.8089	0.8323
KMGCN-NoVisual	0.8621	0.852	0.8166	0.8339
KMGCN-NoKD	0.8690	0.8675	0.8690	0.8677
KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

PHEME

Experiments

Case Study

(a) **The pilot actually took a selfie in the air.**

(b) **Judge Calls For US Marshals and FBI To Arrest Congress and Obama.**

can detect several semantic objects, such as "people", "airplane" and “clouds"

utilize knowledge distillation to obtain knowledge concepts about "judge", "FBI", "Congress", and "Obama" as background knowledge

- Added into textual content of the post to provides complementary information for fake news detection.

Conclusion

- Propose a novel Knowledge-driven Multimodal Graph Convolutional Networks (KMGCN).
  - Conduct the knowledge distillation from a knowledge graph and get a set of concepts to complement the semantic representation of short texts of posts.
  - Model the multimodal content as graphs including post text, visual information to capture non-consecutive semantic features and make use of external additional knowledge information.
- In the future, the authors plan to use a more efficient way to extract the visual information to help our model recognize fake news.

Comments
of KMGCN

- Utilize the knowledge graph as complement information for detection.
  - May can deal with newly emergent event.
- Use conduct post as graph to capture non-consecutive and long-range semantic relations.
  - Not sure PMI score can handle
- Use object detection to capture visual information.
  - Cons: detected object are too general to can't provide full information.